

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2002, 08:58:58 ; Search time 4437.29 Seconds
(without alignments)
17388.122 Million cell updates/sec

Title: US-09-917-384-2
Perfect score: 3687
Sequence: 1 atggagcgaaccacaacatc.....tcagctgcacagccagctga 3687

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_ov:*

5: gb_ov:*

6: gb_ov:*

7: gb_ov:*

8: gb_ov:*

9: gb_ov:*

10: gb_ov:*

11: gb_ov:*

12: gb_ov:*

13: gb_ov:*

14: gb_ov:*

15: gb_ov:*

16: gb_ov:*

17: gb_ov:*

18: gb_ov:*

19: gb_ov:*

20: gb_ov:*

21: gb_ov:*

22: gb_ov:*

23: gb_ov:*

24: gb_ov:*

25: gb_ov:*

26: gb_ov:*

27: gb_ov:*

28: gb_ov:*

29: gb_ov:*

30: gb_ov:*

31: gb_ov:*

32: gb_ov:*

33: gb_ov:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Match Length DB ID Description

1	483.6	13.1	3240	1	CFICELASE	L25809 Cellulomona
2	427	11.6	3004	1	ACU33212	I23212 Acidothermu
3	427	11.6	3004	6	I23701	I23701 Sequence 6
4	339.4	9.2	2535	1	AB078006	AB078006 Streptomy
5	324	8.8	3404	6	I69280	I69280 Sequence 1
6	324	8.8	3503	1	TFU18978	U18978 Thermomonas
7	320	8.7	41906	1	SC567	AL031515 Streptomy
8	314.6	8.5	1269	6	I69283	I69283 Sequence 4
9	314.6	8.5	1365	1	AF268074	AF268074 Thermomon
10	293	7.9	213050	1	AL646079	AL646079 Ralstonia
11	275.6	7.5	1231	1	AF172344	AF172344 Cellulomo
12	264.2	7.2	11541	1	AE003960	AE003960 xyloella f
13	181.6	4.9	3227	1	AF200304	AF200304 Caldicbac
14	171	4.6	4367	1	AF163837	AF163837 Caldicbac
15	134.2	3.6	11707	1	AF078737	AF078737 Caldicell
16	134.2	3.5	11707	6	E35099	E35099 Truncated c
17	130.6	3.5	5437	1	CASR69XN2	AF036924 Caldicell
18	117.2	3.2	165849	2	AC096270	AC096270 Rattus no
19	116.4	3.2	2977	6	A28170	A28170 B. lautus st
20	114.8	3.1	2831	1	BACCELAA	M76588 Bacillus la
21	114.6	3.1	3262	1	AF078038S1	AF078038 Caldicell
22	113.6	3.1	3776	1	CFIB14C	L38827 Cellulomona
23	110	3.0	131995	14	OPU75930	U75930 Orxyia pseu
24	108.4	2.9	749	14	NEOPRT2	D13755 Multicapsid
25	107	2.9	18365	2	AL645600	AL645600 Mus muscu
26	104	2.8	13669	14	TFV1TP	X14855 Thermoprote
27	101.2	2.7	3480	1	CFICEND	L02544 Cellulomona
28	100.8	2.7	1621	1	THEFE2AA	M73321 Thermomonas
29	100.2	2.7	4034	14	TFV1TP	X14717 Thermoprote
30	98.8	2.7	6005	1	AF078038S5	AF078042 Caldicell
31	98.8	2.7	6416	6	E35100	E35100 Truncated c
32	98.6	2.7	3701	1	AF172724	AF172724 Caulobact
33	98.6	2.7	11283	1	AE005710	AE005710 Caulobact
34	96.4	2.6	5513	1	AT286105	Z86105 A. thermophl
35	96.2	2.6	6510	6	AX251402	AX251402 Sequence
36	96.2	2.6	6510	6	AX251811	AX251811 Sequence
37	96.2	2.6	6510	6	AX345297	AX345297 Sequence
38	96	2.6	5439	1	CDCCELA	L32742 Caldocellum
39	95.2	2.6	1116	1	AF233376	AF233376 Streptomy
40	95.2	2.6	1116	6	AR129926	AR129926 Sequence
41	95.2	2.6	1116	6	AR168360	AR168360 Sequence
42	94.8	2.6	330	6	AR135147	AR135147 Sequence
43	93.8	2.5	4977	1	CDCMANA	L01257 Caldocellum
44	93.4	2.5	786	6	AX339682	AX339682 Sequence
45	93.4	2.5	1877	6	AX339680	AX339680 Sequence

ALIGNMENTS

RESULT 1

LOCUS CFICELASE 3240 bp DNA linear BCT 20-JUL-1994

DEFINITION Cellulomonas fimi cellulase, complete cds.

ACCESSION I25809

VERSION I25809.1 GI:456028

KEYWORDS cellulase.

SOURCE Cellulomonas fimi (library: ATCC 484) DNA.

ORGANISM Cellulomonas fimi

REFERENCE 1 (bases 1 to 3240)

AUTHORS Melnick, A., Gilkes, N.R., Kwan, E., Kilburn, D.G., Warren, R.A. and Miller, R.C., Jr.

TITLE Cellulohydrolyase A (cbha) from the cellulolytic bacterium Cellulomonas fimi is a beta-1,4-exocellulohydrolyase analogous to Trichoderma reesei CBH II

JOURNAL Mol. Microbiol. 12 (3), 413-422 (1994)

MEDLINE 94344030

FEATURES

location/Qualifiers

1..3240

/organism="Cellulomonas fimi"

/db_xref="taxon:1708"

```

CDS      /tissue_lib="ATCC 484"
          241..2859

```

```

/codon_start=1
/translation=1
/product="cellulase"
/protein_id="AAC36898.1"
/db_xref="GI:456029"

```

/transilaioyyn--mstlcknagvrrvraayadpplaaakrtvavaavpplttlatsasaaavp
 vdnryaasaooyvnpmpstlcknagvrrvraayadpplaaakrtvavaavpplttlatsasaaavp
 adnglglfhlidnvaapqokaagvplvntvlylplqgdcalaslnselpatmditlarkg
 seeyidpdladliddesvstiatltpdsplntlispcaqaaipyrgovayaldlarkg
 klayhlpwvnyldidgshgwomnagppltpfaleavastagafisvdayantlarkg
 plleplldlsdlsllntmptlksfeyemndpeiltryahmrllyaaefpssicmldv
 tsrnagmcgprpistatnptdnyvaayvnaavndvbrhrnagmcnclagigfepreapscv
 aashldafwlrkpgesdastldipdocrkrdpctpfvpsklntnolctatnnapl
 agovefoeyvplvknvnpvlggtfpedvavpplvgtlagaatnsvslmststnldn
 vaavnygvogkgtlvgtatnsvgtvlgplapaystpvakkaatnsvsaaaaatnldn
 oascvntvtpaspadltaglgtvntvplpsmdnagmcsvacyevlrcrtvnyvgtl
 tasyvntvgtltagvtsfsebradvagnstasasaavntogvntvlatvsvplglt
 gctgttsvsvlvaaynaandnagmcvacyevnvgtrvayntvsttyvlglaadpaysf
 tyakadvagvnsasasakrtadtsccvtyatshmsntltagvsvvknngnaglalnl
 gwtlgsfseadgocvswasemsostavtarnapnmtltagvsvvsgfngtngntrn
 tapnraflnlnactlglg"
 5539. 2859.

```

/bound_moiety="cellulose"
BASE COUNT      509 a      1250 c      1085 g      396 t
ORIGIN

```

Query Match	13.1%	Score 483.6;	DB 1;	Length 3240;
Best Local Similarity	58.7%	Pred. No. 8e-38;		
Matches 1002; Conservative	0;	Mismatches 669;	Indels 37;	Gaps 8;

[illegible][illegible]

Db 1957 GACACGACGGCGCCGTCGCTCCGCCG 1984
 RESULT 2
 ACUC3212
 LOCUS ACUC3212 3004 bp DNA linear BCT 19-SEP-1995
 DEFINITION Acidothermus cellulolyticus E I beta-1,4-endoglucanase precursor
 gene, complete cds.
 ACCESSION U33212
 VERSION U33212.1 GI:988299
 KEYWORDS
 SOURCE
 ORGANISM
 .
 Acidothermus cellulolyticus.
 Acidothermus cellulolyticus
 Bacteria: Firmicutes: Actinobacteria: Actinobacteridae;
 Actinomycetales; Frankineae; Acidothermaceae; Acidothermus.
 1 (bases 1 to 3004)
 Laymon, R. A., Himmel, M. E. and Thomas, S. R.
 Direct Submission
 Submitted (04-AUG-1995) Steven R. Thomas, Applied Biological
 Sciences, National Renewable Energy Laboratory, 1617 Cole Blvd.,
 Golden, CO 80401, USA
 FEATURES
 SOURCE
 Location/Qualifiers
 1..3004

stem_loop	710..725	/db_xref="taxon:28049"	/feature="Repeat" note="three direct repeats of ACGGGGA"
RBS	762..769		
repeat_region	781..817		
	/note="three direct repeats of ACGGGGA"		
	/rpt_type=direct		
	/rpt_unit=781..788		
	824..2512		
CDS	/codon_start=1		
	/transl_table=1		
	/product="E I beta-1,4-endoglucanase precursor"		
	/protein_id="AAV5477.1"		
	/db_xref="GI:988300"		
	/translation="MPRALRVPGSRVRLRGVYVAVLATVAALANLAVPPARAAGG		
	GYWTSRGREILDANNVPRVIRAGINWPGFETGCVNYVHGLMSRSDYRSMIDQISLQNT		
	RLRPSDILKRGTPMNSINFYQNMDDQGLSLDVMKIVAYACIGIRITLDRRRLP		
	CSGSGALANLVSSVEATWISDLQALAQYKGNPTVGVFDLNEHPDPACWCGDPSII		
	WRLLAERAGNAVLSVNPNLLIFEGVQSYNDSDYMWGSGNLQAGQYPVVLVPPNRLVY		
	SAHAYVSYVPTQWFSPTPEFPNNAPGINKMGYLEFNQNIAPVILGEGTGLQSTDQ		
	TKTLTIQYLRKPTAOYGADSFQMTGFMNSNPPSGTGGLIKDMQTVDFVKKGYLAPI		
	SSIDPVGASASPSSSQSPSPSPSPSPASRRPTPTPTPTASPTPLTPTPTPTPTPT		
	ASPTSPPTAASGARTASTQYVNSDMGNEFVTAVTVNSGSVATKTVVSWTFGGQTTT		
	TNSNAAPVTONGOSVTRAKMSYNVNIOPGQNTTGGFQASYSYGNSNAAPVACAS"		
sig_peptide	824..946		
mat_peptide	947..2509		
	/EC_number="3.2.1.4"		
	/note="cellulase"		
	/function="cleaves beta-1,4-glycosidic linkages of		
	cellulose"		
	/product="E I beta-1,4-endoglucanase"		
BASE COUNT	2514..2560		
ORIGIN	604 a 931 c 938 g 531 t		
Query Match	11.6%	Score 427;	DB 1; Length 3004;
Best Local Similarity	99.8%	Pred. No. 2e-32;	
Matches 438; Conservative	0; Mismatches	0; Indels	1; Gaps 1;

QY	61	CCGCGACATCTCAAAACGAGCTGCGAACCAGCGCGTCTCCCGCGGCGGATGACATCGAACC	120
Db	2637	CCCGCATCTCAAAACGCGCTCGAACCAGCGTCTCCCGCGGCGGATGACATCGAACC	2686
QY	2567	ATGGAGCCAAACCAACCAATCCGACGGAACTCCAGGTACCAAGAGGAACGACACGAAAG	2626
Db	1	ATGGAGCCAAACCAACCAATCCGACGGAACTCCAGTACCAAGAGGAACGACAGAAAG	60

Oy	121	tcacatcgccgcgtgaggatcagatctctgcattccatcgccgcgcgcgaacgaatctgcacattcc	180
Db	2687	ttccatctgagccgcgtgcacatgagcatctgcattcgcgcgcacacgtgcacattccc	2746
Oy	181	tatgcggagcgaacctcttcgtcaaccgttaactggcgcaagaagtaacagcgaacg	240
Db	2747	tatggggagagcacccttcttctgtcaaccctgtaactggcgccaagaagtaacagagccaa	CG 2805
Oy	241	gcgaaccgagaccgaatgcacctctgcgcgcgcgaagaatgcgcgtcttcacatatctcagc	300
Db	2806	gcgaacgcagaccgaatgcccactctgcgcgcgcgaagaagaatggcgcgttccacatatctgcacg	2865
Oy	301	gcccgtctgtagtaggacgcgcacgtctgcgatcaacggcgtctcaacgcgcgcgcgcgttcgacg	360
Db	2866	gccgcgtctggatggagaccgcgcacgtctgcgatcaacggcgcgtctcaacgcgcgcgcgcgttcgacg	2925
Oy	361	acatatctggagacgcgcgcctctcccaagcagcgaaggaaacacccctcgaagtcatctgaatt	420
Db	2926	acattttcggagaccgcgcctctcccaagcagcgaaggaaacacccctcgaagtcatctgaatt	2985
Oy	421	gtcacctacgattctgcgcg	439
Db	2986	gtcatctacgattctgcgcg	3004

[illegible]

Query Match	Similarity	11.6%	Score 427	DB 6	Length 3004
Best Local	Similarity	99.8%	Pred. No. 2e-32		
Matches	438	Conservative	0	Mismatches	0
				Indels	1
				Gaps	1
Qy	1	atggagcgaaacccaacatctcgacgagacatgataccagagagagacacacgaatg	60		
Db	2567	ATGGAGCGAACCCACCAATCCGGAGCGAATCTCAAGTACAGAGAGAAACGACGAATG	2626		
Qy	61	cccgccatctcaanaacgagctgcgagccgagcgtctctgcgcgagggcggttagacatcgaccc	120		
Db	2627	CCCGCCATCTCAAAACGGCTGCGAGCCGGCGTCTCGCCGGGGCGGTGAGCATCCAGCC	2686		
Qy	121	tccatctgycgcgctgycgagatcagacatctctgcatctgcgcgcgagacgaactgacaatccc	180		
Db	2687	TCCATCTGTCGCGCTGTGGCATGCAAGCATCTCTGCCATCGCGCCGACCCAGCTCACCAATCCC	2746		
Qy	181	tatgctggagagcagactctctctgcaaccctgatactgagcgcaagaagtacaagaagaaagcg	240		
Db	2747	TATGCGGAGAGCACTTCTCTGTCAAACCCGTTCTGGCGCAAGAAGTACAGAGCCAA-CG	2805		
Qy	241	gcgaaccagaccaaatyccactctgcgaagcgaanaatgcgctcttccacatatctcgagc	300		
Db	2806	GCGAACCAAGACCAATGCACTCTCCACAGGAAATAAGCGGTGTTCACATATTTCGACG	2865		
Qy	301	gcccgtctgtagagacagcatcgctctgcatcaagcgcttcaaacggcgagaccggcttgacg	360		
Db	2866	GCCGCTCTGATGAGACCGGATCGCTCTGATTCACAGGCGCTCAACAGCGCGACCCGCGCTTGACG	2925		

FEATURES	source
LOCUS	AB078006
DEFINITION	2535 bp DNA linear BCT 19-JAN-2002
ACCESSION	Streptomyces sp. M23 gene for exoglucanase CBHT, complete cds.
VERSION	AB078006
KEYWORDS	AB078006.1 GI:18250950
SOURCE	Streptomyces sp. M23 (strain:M23) DNA, clone_11b:pUC 118
ORGANISM	Streptomyces sp. M23 genomic DNA.
REFERENCE	Streptomyces sp. M23
AUTHORS	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
TITLE	Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomyces.
REFERENCE	1 (sites)
AUTHORS	Park,C., Kawaguchi,T., Sumitani,J. and Arai,M.
TITLE	Cloning and sequencing of an exoglucanase gene from Streptomyces
REFERENCE	sp. M23, and its expression of Streptomyces lividans
AUTHORS	Unpublished
TITLE	2 (bases 1 to 2535)
REFERENCE	Park,C., Kawaguchi,T., Sumitani,J. and Arai,M.
AUTHORS	Direct Submission
TITLE	Submitted (15-JAN-2002) Motoko Arai, Osaka Prefecture University,
REFERENCE	Graduate School of Agriculture and Biological Sciences, Gakuen-sho
AUTHORS	1-1, Sakai, Osaka 599-8531, Japan
TITLE	(E-mail:motoo@biochem.osaka-fu-u.ac.jp,
REFERENCE	URL:http://www.biochem.osaka-fu-u.ac.jp/AM/MAIN-J.html,
AUTHORS	Tel:81-72-254-9465(ex.2476), Fax:81-72-254-9921)
TITLE	Location/Qualifiers
REFERENCE	1. .2535
AUTHORS	/organism="Streptomyces sp. M23"
TITLE	/strain="M23"
REFERENCE	/db_xref="taxon:184457"
AUTHORS	/clone_11b="pUC 118 Streptomyces sp. M23 genomic DNA"
TITLE	/country="Japan: Osaka, MIZUMA"
REFERENCE	224. .228
AUTHORS	/evidence=not_experimental
TITLE	241. .1602
REFERENCE	/gene="cbht"
AUTHORS	/gene="cbht"
TITLE	/evidence=experimental
REFERENCE	241. .1602
AUTHORS	/gene="cbht"
TITLE	/standard_name="cellulohydrolase"
REFERENCE	/EC_number="3.2.1.91"
AUTHORS	/codon_start=1
TITLE	/transl_table=1
REFERENCE	/evidence=experimental
AUTHORS	/product="exoglucanase CBHT"
TITLE	/protein_18="BAB83928.1"
REFERENCE	/db_xref="GI:18250951"
AUTHORS	/translation="MSRSRTMLALTLAAGSMTLAAGPASAGPAAPARVNDPVY"
TITLE	GATWYVEMSAALAEPEGCDRVADQPAVWLDRATIEGVADKMGLEHDEALQOQ
REFERENCE	GSSEYVQVLYLDLPEGRCAALASNGELGPELDRIKSEYIDPLADILDSKYEGR
AUTHORS	VTVLEPSLPLVNLNAGTDTTTEACTMKRNGKYEKYSALSKGLIPVYNTID
TITLE	AHHRGLMDITLGPVSVEFTKVAATNGASVDVAGFAVNTANTSPTEVPIVTSDFVNA
REFERENCE	GQVROSKWMDWMDQYVDEQSYAQLRNEVAAGNSDVGVIIDTSRNMGSGSDRSRG
AUTHORS	GQPSVDAVYDGSRIDRRVHVGNMNCOSGAGLGERPTAAPASGIDAYVWIKPPESDQ
TITLE	NSAIVYDNDEGFPQWCDPSYQGNARNGYNPSGALPDAPLSQWFSAGFRELQNAVY
REFERENCE	PLS"
AUTHORS	1629. .1695
TITLE	/evidence=not_experimental
REFERENCE	386 a 1031 c 823 g 295 t
AUTHORS	terminator
TITLE	ORIGIN

[illegible]

Oy	1307	gcgaccagcgttaacgcagcgctgcgcggagtacttccaaac-----gcgtaacga	1357
Db	2132	GGGAGCCCACTTATCGAGGCAACGCCCGCCAAACGCACACCCTCGGTGGCGTCGCCA	2191
Oy	1358	actcccgatcgccgcgcagtttgttcgccggcgccagtttgacaacctgttcgcgaacgac	1417
Db	2192	ACGCCCATCTCCGGCAGACTGCTGTTCTGTCGCCAGTTTCGGAGAGTGCTGGCCAACGCT	2251
Oy	1418	ggccagggtgtgcgcagctgcagcacagctcgagccgcgcgccttcgcgcgcagattccg	1477
Db	2252	ACCCGCTCTGTAAAGCGGAGTAGGCAACGGCTGACAGCCTCAACGAGGAACGTATCAG	2311
Oy	1478	cttcgcgcagttcgagcccgagtcgcgaacccgagagagctgcacatcgccg	1527
Db	2312	CACCTCTTA6CGGAGACGGCGCCGTCACATCCCCGThGGGCGGGCGCTG	2361
RESULT	6		
LOCUS	TFU18978	3503 bp	DNA linear BCT 13-NOV-2001
DEFINITION	Thermomonospora fusca beta-1,4-exocellulase precursor gene,		
ACCESSION	U18978		
VERSION	U18978.1	GI:664822	
KEYWORDS	.		
SOURCE	Thermobifida fusca.		
ORGANISM	Thermobifida fusca. Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptosporangineae; Thermomonosporaceae; Thermobifida. 1 (bases 1 to 3503) *		
REFERENCE	Zhang,S., lao,G. and Wilson,D.B. Characterization of a Thermomonospora fusca exocellulase JOURNAL Biochemistry 34 (10), 3386-3395 (1995)		
AUTHORS	PUBMED 7880834		
MEDLINE	2 (bases 1 to 3503) Zhang, S.		
REFERENCE	Direct Submission Submitted (21-DIC-1994) Sheng Zhang, Biochemistry, Molecular & Cell Biology, Cornell University, 460 Biotechnology Building, Ithaca, NY 14853, USA		
JOURNAL			
TITLE			
AUTHORS			
FEATURES	Location/Qualifiers		
source	1..3503		
	/organism="Thermobifida fusca"		
	/strain="YX"		
	/db_xref="taxon:2021"		
	/clone="ps24; ps26"		
	249..261		
protein_bind	/gene="E3"		
	/bound_moiety="regulatory protein"		
gene	249..2506		
	/gene="E3"		
	/gene="E3"		
5' UTR	262..574		
	/gene="E3"		
stem_loop	349..383		
	/gene="E3"		
protein_bind	350..363		
	/gene="E3"		
	/bound_moiety="regulatory protein"		
stem_loop	467..504		
	/gene="E3"		
RBS	560..564		
	/gene="E3"		
CDS	573..2365		
	/gene="E3"		
	/EC_number="3.2.1.91"		
	/note="Cel C precursor"		
	/codon_start=1		
	/transl_table=11		
	/product="beta-1,4-exocellulase precursor"		
	/protein_id="AAA62211.1"		

[illegible]

[illegible]

	gene	3291..3707 /gene="SC5C7.04" 3291..3707 /gene="SC5C7.04" /note="SC5C7.04, possible lyase, similar to e.g. LGUN_HAEIN lactoylglutathione lyase (EC 4.4.1.5) (135 aa), fasta scores; opt:124 z-score: 215.2 E(): 9.7e-05, 25.08 identity in 131 aa overlap"
CDS		/transl_start=1 /product="putative lyase" /protein_id="CAA20616.1" /db_xref="GI:3559994" /db_xref="SP REMBL:O86701"
RBS		/translation="MDISLSOCETIAVDHHDRALEFYADVLGLEVRNDVGYEGMRWTLGSEAPDVIVILEPPLADPNASADKCMAMDRLAKRILITLTDVDTFFEHVRNAA GAETLEQRPDPDQPYGVCRDCAFRDPAFGNNYRLRPRKG"
gene		3978..3981 /note="possible RBS upstream of SC5C7.05"
CDS		3994..4794 /gene="SC5C7.05" 3994..4794 /gene="SC5C7.05" /note="SC5C7.05, probable RNA polymerase sigma factor, len: 266 aa; similar to S. coelicolor sigma factors of the sigB subfamily e.g. TR_052313 (EMBL:AF036131) putative RNA polymerase secondary sigma factor sigG (263 aa), fasta scores; opt:1258 z-score: 1540.5 E(): 0, 72.9% identity in 251 aa overlap, and to RPOF_STRAU RNA polymerase sigma-F factor (297 aa), fasta scores; opt: 713 z-score: 858.3 E(): 0, 42.6% identity in 312 aa overlap. Possible alternative start 216 bp upstream, not supported by Frapneplot. Contains helix-turn-helix motif at aa 228-249 (Frame 2104, +6.35 SD) and P1am match to entry PF00140 sigma70, Sigma-70 factors, score 210.40, E-value 2.8e-59 /codon_start=1 /transl_table=11 /product="putative RNA polymerase sigma factor" /protein_id="CAA20617.1" /db_xref="GI:3559995"
Query Match		8.7%; Score 320; DB 1; Length 41906;
Best Local Similarity		55.5%; Pred. No. 1.3e-22;
Matches 773; Conservative		0; Mismatches 565; Indels 54; Gaps .6;
OY	73	aaagcgctgcgagcgcgcgctcctgcgcggcggtgatgcacgcactcatgctccg 132
Db	38525 AACGGATCCCGGCCACCTTCACCTCAAGCGCGTGACCTGCACGTGACGGTGGACC	38584
OY	133 ctgcgatgcagcaatccctgcacatcgccgcgcgcgcgcgcgaatccctatgcgagacg 192	
Db	38585 GACCGGACCGAACCCGCCACCGACCCGACGACCCGCCACCGACCCACCGCGCCGGC	38644
OY	193 acctcttcgtcaaccgylactyggcgcaagaagtacagagcgaagcgcgaaccagac 252	
Db	38645 GACCGGGTGGACAMACCTTAGACGAGCGGCCCAAGTCTACGTGAACCCGAGTGGCCGC	38704
OY	253 aatgcacctctgcgcgcggaataatgcgcttccattccaatatcgaagcgcttgatg 312	
Db	38705 AACGCCGCCGCGAACC GGCGGCGCACCGCATCGCGGACGACGACCACCGTGTGCTC	38764
OY	313 gaccgcacgcgtcgatcaacgcgcgtcaaacgcgcgaaccgcgccttgaacaatalctgac 372	
Db	38765 GACCGGATCGCGCGCATGAGAGGGCCCCGAAAGCGCGCATGCGTCTGCGGACCACTGGAC	38824
OY	373 gccgcctctcccaagcaggaaccaacccttgtagtcaatcatgatgtcatctacg 432	
Db	38825 GCGGCGCTGG---ACACAGAAGGGCTTCGGCGAGATGGTGTCCACTGTATCATCAAC	38881
OY	433 ctgccgcgaagcgaatgcgcgcgctgcgcctcaacgcgcgaactgcgcgctaaggacga 492	
Db	38882 CTGCCGGAGCGAGATGTTCCGCCCTCTCCATTCAACGCGCGAGCTGGGCCGACGATGC	38941

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
OY 947	acgaaagcgcagctacgcagttgactgtactcgtcgcgctcgttcgctgcgttcgttcctcaaga														
Db 791	ACGAGCTCTCCTCTCCAGCAGCTGCCTCAGCCCTCATTCGCCAAGGCGTTTCCGCTCG														
OY 1007	gcattcgagctgctacacacctaagcaacagcttggtgtgtcttcggaacgaacaaag														
Db 851	ACATCGGATGATGTCATTCGACACCTCCCGCACMGCGTGGGGTGGGCCGAAACGTCGACCG														
OY 1067	gcccgagcagcgcgagcagatgcaaacaccttcgtcaaccagatcgaaagtatgaccttcgac														
Db 911	GACCCAGCTCCTCCACCCAGACCTCAACACCTACGTTGAGAGAGAGCCGATTCGACCCGTA														
OY 1127	agcagcagcgctctgtgtgtgcaaaccaagcaacgctgtgcgctcgcgacgcgcgcgaagca														
Db 971	TCCACCCCGGTAACGTGGTGCAACCAAGGCGGTGGCGGCTCGGCGACCGGCC														
OY 1187	gcccgagcagctctcccggaagcgcacacctcgagcgtatgtctgtatgaagccgcggtgt														
Db 1024	-----ACGGTCAACCCGGGCTCCCGGTGTGACGCTTACGTCGTGGGTGAAACCCCGGGTG														
OY 1247	agtcgagcagcgaacaaagcgtctgcgagcgtatccgacaaacttgcgaagaagtcgaaacccatgt														
Db 1079	AGTCCGACGGCGCCAGCAGAGAGATCCGAAACGACGAGAGGGCAAGGGCTTCGACCGCATGT														
OY 1307	gcgaccgcagctacacgcagcgtctgcggtgtctctacaaac-----gcgttacgga														
Db 1139	GGGACCCGACCTTACCAAGGGCAACGCCCGCAACGCGAACACCCCTCGGGTGGCGCCCA														
OY 1358	actcccgatcgcgcgcaagctgtctccggcgcgagttcgaccagcttcttgcggaacgac														
Db 1199	ACGCCCCCATCTCGGGCACACTGGTCTCTGTGCCAATTTCCGGAGCTGTGCGCAACGCT														
OY 1418	ggccagcgagtg 1428														
Db 1259	ACCCGCCCTCTG 1269														
RESULT 9															
LOCUS	AF268074														
DEFINITION	Thermomonospora fusca cellulase precursor (Ex) gene, complete cds.														
ACCESSION	AF268074														
VERSION	AF268074.1														
KEYWORDS	GI:8489860														
SOURCE															
ORGANISM	Thermobifida fusca.														
REFERENCE	Thermobifida fusca.														
AUTHORS	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;														
TITLE	Fungal; Actinobacteria; Streptococcaceae; Thermomonosporaceae;														
JOURNAL	Thermobifida.														
REFERENCE	1 (bases 1 to 1365)														
AUTHORS	Al.Y.-C. and Wilson,D.B.														
TITLE	Genomic DNA sequence encoding Thermomonospora fusca cellulase with														
JOURNAL	Unpublished														
REFERENCE	2 (bases 1 to 1365)														
AUTHORS	Al.Y.-C. and Wilson,D.B.</														

	gene	1..1365	/db_xref="GI:8498861"
	mat_peptide	94..1362	/translation="MSRPRLALIGAAAALVSAALLAFPSQAANPGEKVDNPFEGAKLVYVPWASAKAAAPGGSVAANEESTAVMLDRIGALIGNSPTTGSGLRDHLEAVRQSGDPLTIQIVLYINLGRDCALALANGELGPDLBYRKEEVIDIPDIIMDPADYENLRVAILTEIDSLPMLVTNNVGNGSGETELCAVMKNQGVVNCVYALRIEGLIPNVANYIDAHHGWICGMDNSGPSVDVFIEPAANSGSTVDYVHFISNTANTYSATVEPYLDVNKTVNGSLIQSKRWVDNMNOYVDELSEFODURLIAGFEEDTDGLDTSIRNMGGNRPETSPSSDTLNITVEDSRIDRIRHPGNMNOAGAGEEPYNPAGVAYVWVKPPESDGASEELPNDEGKGFDRCMDPTYOGNARNCNPSGALPNPINSGHWFSAOFRELLANA YPLY"
		/gene="Ex"	
		/gene="Ex"	
		/product="cellulase"	
BASE COUNT	242 a	512 c	402 g 209 t
ORIGIN			
Query Match	Best Local Similarity	8.5%; Score 314.6; DB 1;	Length 1365;
Matches 732;	Conservative 0;	Mismatches 479;	Indels 60; Gaps 7;
Oy	200	tcgtaaccgctactggcgcgaagaagtacagagccgaagcggcgaaccagaatgccca	259
Db	110	TGCACAMCCCGCTTCAGAGGGCCGCAAGCGTGACGTACGTACGCCGCTGTGTCGGCCGAAGGCCG	169
Oy	260	cctcgcagcagaataatgscgttcgtttccaatatitgacggccgttcatgttgaaaca	319
Db	170	CCGCGTAGCCCGGGCGGTTCCCGGGTCCGCGCMAAGCATCACCGCTGTCTGGCTGGACGCTA	229
Oy	320	tcgctgcgatcaaacggcgltcaacggcggacccggcttacgacatalatitggacgcgcgc	379
Db	230	TGCGGCGCATCGAGGGCAACGACACCAGCCGACACCGGCTCCATGGGCTGGGCGCACACC	289
Oy	380	t-----ctccagcagcagggaaggaaccacctgaagttaatitgatatitcatatcg	430
Db	290	TGAGAGAGCGCGTCGCCGACAGTCCGGGTGGCGAACCCGCTGACCATCCAGGTCGTCACTTACA	349
Oy	431	atcttcgggagcgcgactcgcgcggcgctcgcctccaacggcgaaactgcgcgtacggag	490
Db	350	ACCTCCCGGCCCGCCACTGCGCGCGCTGGCTCCACAGGTGAAGTG-----GGTCCG	403
Oy	491	caggtttgcagacctatgaacgcagtlacatcgatlcgcatttgcaglatcttcgag----ca	547
Db	404	ATGAACCTCGACCGCTACAAAGAAGCGATGACATGACCGCATGGCCGACATCATMTGTGGACT	463
Oy	548	atccgaagtaactccagccttcggagtcgttgacgataltatggaccggaactcgttcccgaag	607
Db	464	TGCGAGACTTACGAGAACCTGGGAGATCGTGGCCATCATGCAATGCATCTCCCTGCCCAACC	523
Oy	608	cggttaccaatatitgagcatlcaaaggtitlgaaaagcgggttgcgcat-----	654
Db	534	TGCTACCAACAGTGCGGGCGGAACGGCGGCACCGACTCTTGCGCTTACATGAAGCAAAG	583
Oy	655	-----tacgagcaagagcatcgagtlacgcgtctoaagaaatlgaacgcacatccgaagcgt	709
Db	584	GCGGCTAGGTCAACGCGTTCGGCTACGGCTTCGCGAACCTGGGCGCAAGATCCCGAAGCTCT	643
Oy	710	acatctacatggagcgcgcaccactcgcgccttggcttgggtggccccaatatitgccagcgat	769
Db	644	ACAACTCATGATGACGCCGCCACACACGGCTGGATTGGCTGGGAACTCAACTTCGGCCCCCT	703
Oy	770	acgttcaggaaggtccaggaaggttcctcaag---cgagcatctgggttcaacggcatcgacg	826
Db	704	CGETTGACATCTTCTTACGAGAGGCCGCAACGGCTCCGGCTCCACCGTGGACTACGTGAC	763
Oy	827	gcttcgtcaccaaacgcggcgaatlaaacgcgctttaagagacgcyttcatlgaacgcgcacc	886
Db	764	GCTTCATCTCCAACACGGGCCAACTACTGCGGCACCTGTGGAGCCGTAACCTGGACGTCAC	823
Oy	887	agcaggttcggcgagacgcggttggagtcggcgaaatttaccagtgaaatccttgcacatc	946

QY	739	tgcttggtggtgcccccaataatgccaagcgagatgcgtacacggaaagtcacaaagtcctctaac	798
Db	883	TGGCTCGGCTGGAGCTCGACCGGGCTCCGGCGGAAGATCTTCGCCGACGTCGCCAAG	942
QY	799	gcgcagcatcggtggtcaacg---gcacgcagcgcttcgtcaccaacagcggaattacacg	855
Db	943	ACGACACACGGGGGCTTCGCGTCATCACCGGCTTCGTCTCGGACGTCGCCAACCACG	1002
QY	856	ccgttgaagagacc	869
Db	1003	CCGCTTGAGAGACC	1016
RESULT	12		
AE003960/c		11541 bp	DNA linear BCT 15-JUN-2001
LOCUS	AE003960		
DEFINITION	Xylella fastidiosa 9a5c, section 106 of 229 of the complete genome.		
ACCESSION	AE003960		
VERSION	AE003960.1		GI:9106242
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
1 (bases 1 to 11541)			
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,			
Alvarenga,R., Alves,L.M., Araya,J.E., Bala,G.S., Baptista,C.S.,			
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,			
Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carer,H.,			
Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,			
Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H.,			
Facinani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,			
Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,			
Gardner,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,			
Ho,P.L., Hohnsels,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.			
and Marino,C.L.			
The genome sequence of the plant pathogen Xylella fastidiosa. The			
Xylella fastidiosa Consortium of the Organization for Nucleotide			
Sequencing and Analysis			
Nature 406 (6792), 151-157 (2000)			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
2 (bases 1 to 11541)			
10910347			
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,			
Alvarenga,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S.,			
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,			
Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,			
Carer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,			
Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,			
Docena,C., El-Dorri,H., Facinani,A.P., Ferreira,A.J.S.,			
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,			
Frohme,M., Furlan,L.R., Gardner,M., Goldman,G.H., Goldman,M.H.S.,			
Gomes,S.L., Gruber,A., Ho,P.L., Hohnsels,J.D., Junqueira,M.L.,			
Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kurmae,E.E., Lalgrei,F.,			
Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,			
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,			
Madeira,M.F.P., Martino,C.L., Marques,M.V., Martins,E.A.L.,			
Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,			
Myaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,			
Nascimento,A.L.T.O., Netto,L.E.S., Nhani,J.A., Nobrega,F.G.,			
Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,			
Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira			
Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,			
de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,			
Sawasaki,H.E., da Silva,A.C.R., da Silva,P.R., da Silva,A.M., Silva			
Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de			
Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,			
Tsubako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,			
Veloso,A.L., Zago,M.A., Zatz,M., Zeldanis,J. and Setubal,J.C.			
Direct Submission			
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and			
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP			
13083-970, Brazil			

FEATURES	Source
Location/Qualifiers	
1..11541	
/organism="Xylella fastidiosa 9a5c"	
/db_xref="taxon:160492"	
/clone="9a5c"	
complement(140..1057)	
/gene="Xf1258"	
complement(140..1057)	
/gene="Xf1258"	
/note="similar to SPIP1666 (percent identity: 35 %/query alignment coverage: 88.5 %/subject alignment coverage: 94.4 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"	
/codon_start=1	
/transl_table=1	
/product="small conductance mechanosensitive ion channel"	
/protein_id="AAP84067.1"	
/db_xref="GI:9106243"	
/translation="MDMTKLISNIVSPATVTTTHGRDWRIDMTHITMLNTAAAIATLIGVLSKRLSNGLHRLIRAHVEITLANELRNIAVALLIIVSVYLORIGVATSLAAVVGATGLAVGLAKDLSNIAAGVAILVGPVSGDHVIAQESITIDEIRFOTRLTFPDORITLTPNSTITTVPLINSTLPTRIETTVGVSYODNKQAOELLKARENPKLIDDPALIVYNNIGESSVDLLLAVTODNPNPAKSELLEQIHNGLEBNGINITYPQRDLHXHDTNRRKIASLLPVADES"	
complement(1058..3457)	
/gene="Xf1259"	
complement(1058..3457)	
/gene="Xf1259"	
/note="similar to SPIP2358 (percent identity: 64 %/query alignment coverage: 99.9 %/subject alignment coverage: 100.8 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"	
/codon_start=1	
/transl_table=1	
/product="phosphoenolpyruvate synthase"	
/protein_id="AAP84068.1"	
/db_xref="GI:9106244"	
/translation="MALNGASVYLNELTLMHELRLTDLARYGKNSIGENITLALGVSPGVATTAAPKAFIAVNNLSKRTYKLAFLDVEDVAKALPAACKETIDMTIDPLOPELEKDIRADYDTLSKNSGNDVAVAASFAEEDLPDPSFAGQOETFLNVGADGVHRIKEFVSLYNDRAIVAVHGHGHEDEFTSAGIOLAWRSDISASGVLTDTDSGFRDVFVTASFGIGEMVVOGAVNPDEFFYKTKTKAGKAPLIRLTGSKLRYVYDSPGRVATIEDPTLRHCFRATSDVDVLEKALQALITEHGRPDITMAKDSGSKFIVQARPEVTSRSHTTQIERPALQKREILTEGSAIQKIGSGISRVVSHEDNHQPGDVLVADMTDPMPEPMKASAVTRNGRTGHAALIAELGVAVGNGALSIQDGPVITSCAEGDGYIYAGKLPERTTDLGNHPAPLKLMMNVANPERAFDGLPNAIGLARLEELIIAHIGVHPKALEYHKKODAPKOKIDAKIAGPPVSVYRLAEGIATLTASVAPHTVIRLSDPKSVEYAHIIIGCPHREOEPMPICGPGASRYNDRSDAPALECKAVIKVNMKGDLNTRMIPFVRLIEGRKIVYIVLRNGALGNGKLVIKCEVFSNALDERLDFDGFTSISNDLTLSGLDSSIVAPLFDERNPAPVKILLSMAIKAAASKGRTYIGICGQPSDHPDLAEWMQEGIESLSINPDTVDTWRLAKITVH"	
3549..4403	
/gene="Xf1260"	
3549..4403	
/gene="Xf1260"	
/note="similar to SPI054458 (percent identity: 41 %/query alignment coverage: 100.7 %/subject alignment coverage: 98.6 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"	
/codon_start=1	
/transl_table=1	
/product="conserved hypothetical protein"	
/protein_id="AAP84069.1"	
/db_xref="GI:9106245"	
/translation="MCRGSIIEGCMSTIRPVYSDGTGTAFTGHSITLTPSGFTFYTRMPTIDDAEKARDAQRITLAASQRYRPRVIVNSCPNPTSTVLASGALMDVFAPIITGLHEHLNLSRSRIGRAHGVDFEYTHRIANMFPALAHGDVAAASVDEAVILVAVSRAGKPTCIYIALHYGIRAAVNPILDEDLNSDOLPLRLPYRKILFGITLPERLIOQIOERPNRVAALDPCKREVAAREMFASERTITLSTHTSIEISISVLVTIGLQREMF"	
4501..4986	
/gene="Xf1261"	

gene

Dd	2946	GAAATCAGCTTCACACTCGGGGCCGCGGGAGCCCTGGCGGCCTGGGGCCGAACAAGCGGAGCACTT	3005
Oy	2065	caaaacccgggtgaataagaagtgcactgctgcgaattaccagcagacctatcgatat	2124
Dd	3006	CAGGCGCGCATCACAAAGAAGACGTGAGCAGCAACTACAAAGAGCGAAGCACTACTCGTAT	3065
Oy	2125	g---ggaagaaacaccacttcacgagcacggcaggaaagtgaaggttgaagttaagtaacggcgtg	2181
Dd	3066	GATCCGACGAAGAACAGATTGTTCGGATTGGAAACC GGGTACGCGTGATCGGAATTGTCAG	3125
Oy	2182	ttagtgctgggggaactgaaccgtccsg	2207
Dd	3126	CTCCTCTGGGGCGCTGAGCCCTTAAGG	3151
RESULT_14			
LOCUS	AF163837	4567 bp DNA linear BCF-08-FEB-2000	
DEFINITION	Calidbacillus cellulovorans multidomain beta-1,4-mannanase precursor (mana) gene, complete cds; and unknown genes.		
ACCESSION	AF163837		
VERSION	AF163837.1	GI:6651325	
KEYWORDS	Calidbacillus cellulovorans. Calidbacillus cellulovorans Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Alicyclobacillus group; Calidbacillus. 1 (bases 1 to 4567) Sunna,A., Gibbs,M.D., Chin,C.W.J., Nelson,P.J. and Bergquist,P.L. A gene encoding a novel multidomain beta-1,4-mannanase from Calidbacillus cellulovorans and action of the recombinant enzyme on kraft pulp		
REFERENCE	Appl. Environ. Microbiol. 66 (2), 664-670 (2000)		
AUTHORS	JOURNAL MEDLINE PUBMED	20120520	
TITLE	10653733	2 (bases 1 to 4567)	
REFERENCE	Sunna,A., Gibbs,M.D. and Bergquist,P.L.		
AUTHORS	Direct Submission		
TITLE	Submitted (29-JUN-1999) Biological Sciences, Macquarie University,		
JOURNAL	North Ryde, New South Wales 2109, Australia		
FEATURES			
Source	Location/Qualifiers		
	1..4567		
	/organism="Calidbacillus cellulovorans"		
	/db_xref="taxon:74586"		
	<1..513		
CDS	/note="orf1"		
	/codon_start=1		
	/transl_table=1		
	/product="unknown"		
	/protein_id="AAF22273.1"		
	/db_xref="GI:6651326"		
	/translation="TPPTPPTSAPRPPPSAGSLVVOYRAADTNAGNDLKPRLTVNRKTSVPPLSELTIRKYWVDGRPOVFNCDAQVCSSNRRGSFYKLSTGRCTADYYIEILFTFGAGSLAAGASSGDIDVRINKNDWTINEADYSIDPKTFADMNVTLTKNGDIIWGVEPR"		
	1..60		
misc_feature	/note="Region: proline-threonine linker"		
	/evidence-not_experimental		
misc_feature	61..510		
	/note="Region: cellulose-binding domain type IIB"		
	/evidence-not_experimental		
gene	620..3412		
	/gene="mana"		
	620..718		
sig_peptide	/gene="mana"		
	/evidence-not_experimental		
	620..3412		
CDS	/gene="mana"		
	/codon_start=1		
	/transl_table=1		
	/product="multidomain beta-1,4-mannanase precursor"		
	/protein_id="AAF22274.1"		

[illegible]

[illegible][illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2002, 09:56:38 ; Search time 378.37 Seconds

(Without alignments)
16730.353 Million cell updates/sec

Title: US-09-917-384-2

Perfect score: 3687
Sequence: 1 atggagcgcaaccacaacatc.....tcagctgcacagccagctga 3687

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	423.8	11.6	3004	17	AA12337	A. cellulolyticus
2	423.8	11.5	3004	21	AA255924	Acidothermus cellu
3	324	8.8	3503	19	AAV07164	Thermomonospora fu
4	322.4	8.7	3404	17	AA15595	Thermomonospora fu
5	314.6	8.5	1269	19	AA15596	Thermomonospora fu
6	251.2	6.8	3033	19	AA194197	Thermomonospora fu
7	134.2	3.6	11706	20	AA55661	DNA sequence encod
8	134.2	3.6	11707	24	AA26525	Active cellulase p
9	116.4	3.2	2977	12	AA013001	Endol gene encodin

10	108.4	2.9	1422	20	AA27948	H. insolens cel6A
11	108.4	2.9	1437	20	AA27958	H. insolens cel6A
12	98.8	2.7	6415	20	AA55662	DNA sequence encod
13	98.8	2.7	6416	24	AA26526	Active cellulase p
14	97.6	2.6	1621	19	AAV07163	Thermomonospora fu
15	96.2	2.6	6510	22	AA54668	Tumour suppressor
16	96.2	2.6	6510	22	ABJ32395	Human immune syste
17	96.2	2.6	6510	24	AA561115	Human gene regulat
18	95.2	2.6	1116	20	AAV72359	Actinomyces sp. 3
19	95.2	2.6	1116	21	AA57029	Actinomyces cell
20	95.2	2.6	1715	21	AA57031	DNA sequence of an
21	95.2	2.6	2942	21	AA48661	L. mycopilius chlt
22	94.8	2.6	390	13	AA21853	Randomising oligon
23	94.8	2.6	390	14	AAQ36859	PCR primer for 5'
24	94.8	2.6	390	22	AA76910	Sequence containin
25	88.6	2.4	4376	21	AA14818	DNA encoding a Cl
26	87.8	2.4	390	13	AAQ2183	Randomising oligon
27	87.8	2.4	390	14	AAQ36859	PCR primer for 5'
28	87.8	2.4	390	22	AA76910	Sequence containin
29	87.8	2.4	4403765	22	AA199683	Mycobacterium tube
30	87.6	2.3	37286	23	AA59522	Humicola insolens
31	86.2	2.3	2409	19	AAV19377	H. insolens cellu
32	86.2	2.3	2409	19	AAV19281	Portion of PAEC-1
33	84.2	2.3	2600	12	AAQ15178	Cellulohydrolase
34	83.6	2.3	1683	6	AAAN50359	Trichoderma reesei
35	82.6	2.2	3241	21	AAAF14880	Streptomyces nous
36	82.6	2.2	65140	22	AA017184	Streptomyces nous
37	82.6	2.2	125401	22	AA017186	Streptomyces nous
38	81.6	2.2	4403765	22	AA199683	Mycobacterium tube
39	81.2	2.2	2029	20	AA55660	DNA sequence encod
40	81.2	2.2	2029	24	AA26568	Active cellulase h
41	81	2.2	4411529	22	AA199682	Mycobacterium tube
42	78.2	2.1	4056	22	AAH75060	Nucleotide sequenc
43	77.6	2.1	2712	20	AA06825	Streptomyces albid
44	77.6	2.1	2712	22	AAH74538	Nucleotide sequenc
45	77.4	2.1	985	19	AAV64548	M. tuberculosis Im

ALIGNMENTS

RESULT 1
AA12337 standard; DNA: 3004 BP.
ID AA12337
XX
AC AA12337;
XX
XX 08-OCT-1996 (first entry)
XX
DE A. cellulolyticus EI endoglucanase gene.
XX
XX EI endoglucanase; cellulase; cellulose; saccharification; ethanol;
KW ss.
XX
XX
OS Acidothermus cellulolyticus.
FH key location/Qualifiers
FT misc_difference 200
FT /*tag- a
FT /*note- "base g at position 200 is shown in Fig 1,
FT /*note- but is missing in the sequence given on
FT /*note- page 22 of the specification"
FT
FT misc_signal 710..725
FT /*tag- b
FT /*note- "putative upstream regulatory sequence"
FT
FT RRS
FT /*tag- c
FT repeat_unit 781..790
FT /*tag- d
FT /*rpl_type- DIRECT
FT repeat_unit 795..804
FT /*tag- e
FT /*rpl_type- DIRECT

```

FT      repeat_unit      810..817
FT      /*tag= f
FT      /rpt_type= DIRECT
FT      CDS
FT      824..2512
FT      /*tag= g
FT      sig_peptide
FT      824..946
FT      /*tag= h
FT      sig_peptide
FT      863..946
FT      /*tag= i
FT      /note= "alternative signal sequence"
FT      mat_peptide
FT      947..2509
FT      /*tag= j
FT      terminator
FT      2514..2560
FT      /*tag= k
XX
XX      W09602551-A1.
XX
XX      01-FEB-1996.
XX
XX      14-JUL-1995; 95WO-US08868.
XX
XX      15-JUL-1994; 94US-0276213.
XX      26-SEP-1989; 89US-0412434.
XX      27-JAN-1992; 92US-0826089.
XX      21-SEP-1993; 93US-0125115.
XX
XX      (MIDE ) MIDWEST RES INST.
XX
XX      Himmel ME, Laymon RA, Thomas SR;
XX
XX      WPI; 1996-105843/11.
XX      P-PSDB; AAR89927.
XX
XX      New isolated DNA encoding endoglucanase - obt'd from Acidothermus
XX      cellulolyticus, used for prodn of the enzyme for use in cellulose
XX      hydrolysis.
XX
XX      Claim 4; Page 22-24; 34pp; English.
XX
XX      The E1 endoglucanase gene (T12337) of Acidothermus cellulolyticus
XX      codes for an enzyme, the mature form of which (R89927) can hydrolyse
XX      cellulose and xylan. The gene was obt'd. by ligating Sau3A fragments
XX      of A. cellulolyticus genomic DNA into EMBL3 lambda phage arms,
XX      packaging the phage DNA, plating on E.coli LE392 agar contg. CMC,
XX      staining with Congo Red, and isolating DNA from positive clones.
XX      Alternative signal sequences were identified (see also R89928-29).
XX      The gene can be incorporated into a vector and used for large-scale,
XX      low-cost prodn. of recombinant E1 endoglucanase, pref. using
XX      Saccharomyces, Zymomonas or E.coli hosts. The enzyme is useful
XX      for the saccharification of cellulosic biomass for fermentation to
XX      ethanol.
XX
XX      Sequence 3004 BP; 604 A; 931 C; 938 G; 531 T; 0 other;
XX
Query Match      11.6%; Score 427; DB 17; Length 3004;
Best Local Similarity 99.8%; Pred. No. 2,4e-65;
Matches 438; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY      1 atggagcgaaacccaacatccgagcgagactcgagctaccagagaggaagcagcaatg 60
DB      2567 atggagcgaaacccaacatccgagcgagactcgagctaccagagaggaagcagcaatg 2626
QY      61 ccgcgcattcaaaacggctgcgagccggcgctcctcgccggggcggtgagcatcgagcc 120
DB      2627 ccgcgcattcaaaacggctgcgagccggcgctcctcgccggggcggtgagcatcgagcc 2686
QY      121 tccatcgcgcgctgctggcatgagatcctgcctcgcgcgcgcgcgcgcgcgcgcgcgcgc 180
DB      2687 tccatcgcgcgctgctggcatgagatcctgcctcgcgcgcgcgcgcgcgcgcgcgcgcgc 2746
QY      181 tatgcgagcgacactctctgtaaccgctactgagcgcgagagtaagagcgagcg 240

```

```

DB      2747 tatgcgagcgacactctctgtaaccgctactgagcgcgagagtaagagcgaa-cg 2805
QY      241 gcgaaccgacccaatgcaactctcgacgagaaatgcgcgctgttccacatattcgacg 300
DB      2806 gcgaaccgacccaatgcaactctcgacgagaaatgcgcgctgttccacatattcgacg 2865
QY      301 gccgcctgcatgagacgcgcatcgctgcgtatcaacggcgctcaacggcgagccgcttcgacg 360
DB      2866 gccgcctgcatgagacgcgcatcgctgcgtatcaacggcgctcaacggcgagccgcttcgacg 2925
QY      361 acatatctgagacgcgcctctccagcagcaggaaccacccctgaattgagatt 420
DB      2926 acatatctgagacgcgcctctccagcagcaggaaccacccctgaattgagatt 2985
QY      421 gtcattcagcatctgcgg 439
DB      2986 gtcattcagcatctgcgg 3004

RESULT 2
AAZ55924
ID      AAZ55924 standard; DNA; 3004 BP.
XX
XX      AAZ55924;
AC
XX      10-APR-2000 (first entry)
XX
XX      Acidothermus cellulolyticus E1 endoglucanase gene.
XX
XX      E1 endoglucanase; cellulose binding domain; CBD; cellulose modification;
XX      beta-1,4-endoglucanase; endocellulase; thermostable; ds.
XX
XX      Acidothermus cellulolyticus.
OS
XX
XX      Key      Location/Qualifiers
XX      CDS      824..2512
XX      sig_peptide 824..946
XX      /*tag= a
XX      /*tag= b
XX      /note= "Encodes putative signal peptide"
XX      sig_peptide 863..946
XX      /*tag= c
XX      /note= "Encodes putative signal peptide (alternative)"
XX      mat_peptide 947..2509
XX      /*tag= e
XX      /product= "Mature E1 endoglucanase"
XX      misc-feature 2204..2509
XX      /*tag= f
XX      /note= "Encodes cellulose binding domain (specifically
XX      claimed)"
XX
XX      CA2226898-A1.
XX
XX      25-SEP-1999.
XX
XX      25-MAR-1998; 98CA-2226898.
XX
XX      25-MAR-1998; 98CA-2226898.
XX
XX      (MIDE ) MIDWEST RES INST.
XX
XX      Laymon RA, Adney WS, Thomas SR, Himmel ME;
XX
XX      WPI; 2000-087663/08.
XX      P-PSDB; AAY69508.
XX
XX      Isolated domains of Acidothermus cellulolyticus E1 endoglucanase useful
XX      for labeling or modifying a cellulose and for purifying or immobilizing
XX      a binding domain fusion protein to cellulose
XX
XX      Claim 12; Fig 1; 85pp; English.
XX
XX      This sequence represents the Acidothermus cellulolyticus E1 endoglucanase

```

CC gene, encoding a beta-1,4-endoglucanase, or endocellulase. The cellulose
 CC binding domain (CBD) of E1 endoglucanase, and nucleotides which encode it
 CC are specifically claimed. The CBD is believed to be roughly wedge-shaped;
 CC as the CBD binds to cellulose, it is thought that the wedge tip is
 CC inserted between the microfibrils of the cellulose fibre, disrupting the
 CC crystalline structure, and making the cellulose linkages more accessible
 CC to the catalytic domain of the E1 endoglucanase. The E1 endoglucanase CBD
 CC is useful in labelling or modifying the surface of cellulose or other
 CC polysaccharides. Such modified cellulose can then be used in textile,
 CC pulp, paper, chemical and pharmaceutical industries. CBDs can be used in
 CC affinity purification of CBD-fusion proteins, and can also be used to
 CC immobilise the CBD-fusion proteins to a cellulose support. CBD-fusion
 CC proteins can be used to modify the chemical or physical properties of a
 CC cellulose or polysaccharide matrix column and to modify (e.g., roughen or
 CC disrupt) a cellulose or polysaccharide fibre. The CBD of the E1
 CC endoglucanase exhibits greater stability at pH 4-8 and has an optimum
 CC temperature for stability of 83 degrees Celsius which is not found in
 CC CBDs from non-thermophilic organisms.

XX Sequence 3004 BP; 605 A; 933 C; 935 G; 531 T; 0 other;

Query Match 11.5%; Score 423.8; DB 21; Length 3004;

Best Local Similarity 99.3%; Pred. No. 8.6e-65; Mismatches 2; Indels 1; Gaps 1;

Matches 436; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

1 atgagcgaacccaacatccgagcgaactgaggtacacagagaggaacgacgaatg 60

2567 atgagcgaacccaacatccgagcgaactgaggtacacagagaggaacgacgaatg 2626

61 cccgcacatcacaacgagcgtgagcgcggtctcgcggggcggtgagcgcgcgc 120

2627 cccgcacatcacaacgagcgtgagcgcggtctcgcggggcggtgagcgcgcgc 2686

121 tcatctgtcgcgtgagcgtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 180

2687 tcatctgtcgcgtgagcgtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2746

181 tatgaggagagcgcgtctctcgtcaacccgactgagcgcgaagaatgacagcgaagc 240

2747 tatgaggagagcgcgtctctcgtcaacccgactgagcgcgaagaatgacagcgaagc 2805

241 ggcgaacgagcgaacatgacatctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300

2806 ggcgaacgagcgaacatgacatctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2865

301 ggcgtctgagtgagc 360

2866 ggcgtctgagtgagc 2925

361 acatctctgagc 420

2926 acatctctgagc 2985

421 gtcattacgagtcgc 439

2986 gtcattacgagtcgc 3004

RESULT 3

AAV07164

ID AAV07164 standard; DNA; 3503 BP.

XX AAV07164;

AC 11-SEP-1998 (first entry)

XX Thermomonospora fusca cellulase E3 gene.

DE cellulase; E3; recombinant plant; cellulose-degrading enzyme;

XX fermentation; brewing; wine-making; silage production; textile;

KM pulping; paper-making; industry; livestock feed; ss.

XX

This appears to be a match
in 3' UTR region.

OS Thermomonospora fusca.
 XX
 FH Key Location/Qualifiers
 FT CDS 575..2365
 FT /tag- a
 FT /product- cellulase E3
 FT mat.peptide 689..2365
 FT /tag- b
 FT /note- "mature cellulase E3"

XX MO9816651-A1.

XX 23-APR-1998.

XX 13-OCT-1997; 97MO-US18284.

XX 26-JUN-1997; 97US-0883495.

XX 17-OCT-1996; 96US-0028718.

XX (WISC) WISCONSIN ALUMNI RES FOUND.

XX Austin-Phillips S, Burgess RR, German TL, Ziegelhofer T;

XX WPI, 1998-251294/22.

XX Recombinant plant containing cellulose-degrading enzyme - used in

XX fermentation processes such as brewing and wine-making

XX Disclosure; Pages 33-36; 52pp; English.

XX The sequence is that encoding cellulase E3 which can be used in

XX the transformation of genetically recombinant plants. Such plants

XX can be used for producing cellulose-degrading enzymes. The

XX cellulose-degrading enzymes produced by the plant can be used

XX to increase production of silage, thereby ensiling the plant.

XX The cellulases produced by the plants can be used in fermentation

XX processes such as brewing and wine-making. They may also be used

XX to increase the extent of fermentation of ensiled matter. The

XX plants may be fed directly to livestock to aid in the digestion

XX of lignocellulosic substrates. The cellulases may also be used

XX in the textile, pulping, and paper-making industries.

XX Sequence 3503 BP; 633 A; 1276 C; 1028 G; 566 T; 0 other;

Query Match 8.8%; Score 324; DB 19; Length 3503;

Best Local Similarity 56.6%; Pred. No. 1.7e-47;

Matches 775; Conservative 0; Mismatches 535; Indels 60; Gaps 7;

200 tcgtcaacccgctactggtgagcgaagaatgacagagcgaacgacgaatgcca 259

1110 tcgaacacccgctcgaagggtgagcgaagtgtgacgacccggtctgtgccaaggcgc 1169

260 ctctcgaacgaagaatgagcgtctgttcacatatctgacgagccgtctgtgacccga 319

1170 ccgctgagccggcggtcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1229

320 tcgctgagcgaagcgtgacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 379

1230 tcggcgagcgtgagcgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1289

380 t-----ctccacgacgacggaacacacccctgagtcattgagtcatacag 430

1290 tggagagagc 1349

431 atctgcggagc 490

1350 acctgcggc 1403

491 caggttgacgacactgaaacgcagacacacacacacacacacacacacacacacacac 549

1404 atgaactcgc 1463

CC may be involved in regulation of expression. The gene may be cloned and expressed in *E. coli* or *Streptomyces lividans*. The gene product, or a catalytically active polypeptide obtained by pepsin digestion, hydrolyses cellulose or chitosan at pH 5-11 and 40-70 °C and has significant activity at over 60 deg C. E3 has higher stability to proteolysis in culture supernatant than T. fusca cellulases-E2 and -E5, and shows strong synergistic activity when combined with other cellulases, cellobiohydrolase and optionally beta-glucosidase. The cellulase and mixtures may be used e.g. in cellulose saccharification for ethanol production, fruit juice clarification, as a surfactant additive, or in delinking or refining of recycled paper.

Sequence 3404 BP; 619 A; 1239 C; 1004 G; 542 T; 0 other;

Query Match	8.78;	Score 322.4;	DB 17;	Length 3404;
Best Local Similarly	56.58;	Pred. No. 3.2e-47;		
Matches 774; Conservative	0;	Mismatches 536;	Indels 60;	Gaps 7

[illegible]

Db	1784	acgagctctcccttgcctccagagacttcgctcagggccctctgactcgcacaaaggctctccgctcg	1843
Oy	1007	gcatcgagatctcatctatcgaaaccttcaagcaagcggttggtggtgtgtccgaaagcaaacag	1066
Db	1844	acatcgtgatctcatctacgaaacctcccgcaagcgctgggtgggtccggaacgcgtccgacag	1903
Oy	1067	gcccgagacccgcgcagcagatgttcaacaaccttcgtccaacagtcgaaagatttgaccttcg	1126
Db	1904	gaccgagctctctccacccgacctcaaaccttactgttagcgagagccgtatcgcacgcgcgta	1963
Oy	1127	agcacgcgcgcctctgtgtgtcaaccagaagctgtgggtgcctcgtgccagccgcgcagca	1186
Db	1964	lccaccccggttaactgtgtgcaaccagagccggtgtgggtccctcgtgcgagc-----	2011
Oy	1187	gcccgagagattcccgcaaacgcgcacctcgcagcgatgtcttgatccagcgcgcgggtg	1246
Db	2012	ggcccaacggttcaaccgcgctcccggtgtgttagcgcttactcgtgtgggttagaagcccggtg	2071
Oy	1247	agtcgagacgcagcaagcgctgtgcgagcgtatccgacaactgtgcaagaagctcgaccacca	1306
Db	2072	agtcgcagcggcgacgagggagatcccgcaagcagagggcgcaagggtcttcgcacgcgatg	2133
Oy	1307	gcgacgcgagctacagacgctcgtacgggtacttgaccac-----gcgttacgga	1357
Db	2132	gcgaccccgaccttacaagggcaagcccgcaacgccaacccctcgggtgcgtgcgcga	2191
Oy	1358	actcccgatcgcgcggcacaatgtttcccgcgcgagtttagcaacgctgtgcggaagcgac	1417
Db	2192	acgcgcccatctccgggcactgtgtctctgcacatcttcgcgagctgtcgtgcacacgct	2251
Oy	1418	ggccagcgggtgcgcagctcgaccagctcgagccgcgcgcctccgcgcgcgcgagctccg	1477
Db	2252	accgcgcctctgtaaagcgaggttagcgaaagcgctgtgacagctcccaacgaggaactgac	2311
Oy	1478	cttcgcgagtcgcagaccgcgagtcgcagcccgagcagctgcgcacatcgccg	1527
Db	2312	cacctctcagcgcgagagcgcgcgcctccactcccgctggcgcgcgccg	2361
RESULT 5			
ID	AAIT15596	standard; DNA; 1269 BP.	
AC	AAIT15596;		
DT	06-APR-1996	(first entry)	
DE	Thermostable cellulase-E3 catalytic domain gene.		
KW	Cellulase-E3; Thermomonospora fusca; thermostable; papain;		
KW	catalytic domain; Escherichia coli; cloning; Streptomyces lividans;		
KW	cellulose; chitosan; glucose resistance; synergism;		
KW	cellulohydrolase; beta-glucosidase; saccharification;		
KW	surfactant additive; paper recycling; delinking; paper refining; ds.		
OS	Thermomonospora fusca strain YK36.		
PN	W09600281-A1.		
PD	04-JAN-1996.		
PF	23-JUN-1995; 95WO-US09069.		
PR	24-JUN-1994; 94US-0265429.		
PA	(CORR) CORNELL RES FOUND INC.		
PI	Walker LP, Wilson DB, Zhang S;		
DR	WPI; 1996-068865/07.		
XX	P-PSDB; AAR90716.		

Cellulase from *Thermomonospora fusca* - is thermostable and is useful in a variety of industrial applications e.g. clarification of fruit juices, fabric softening, etc

PS Claim 1; Page 38-40; 53pp; English.

CC The sequence encodes a catalytically active domain of thermostable
CC cellulase-E3 from *Thermomonospora fusca*, which is isolated by
CC papain cleavage of the full-length protein (AA830715, encoded by
CC AA715595). The binding domain and linker region coding sequences
CC have been removed from the full-length gene. The gene may be
CC cloned and expressed in *Escherichia coli* or *Streptomyces lividans*.
CC The domain retains full activity, but shows reduced cellulose
CC binding. The gene product, or the corresponding full-length
CC cellulase, hydrolyses cellulose or chitosan at pH 5-11 and 40-70
CC deg C, and has significant activity at over 60 deg C. E3 has
CC higher stability to proteolysis in culture supernatant than T.
CC fusca cellulases-E2 and -E5, and shows strong synergistic activity
CC when combined with other cellulases, cellobiohydrolase and
CC optionally beta-glucosidase. The cellulase and mixtures may be
CC used e.g. in cellulose saccharification for ethanol production,
CC fruit juice clarification, as a surfactant additive, or in delinking
CC or refining of recycled paper.

Sequence 1269 BP; 234 A; 474 C; 371 G; 190 T; 0 other;

Query Match	8.58;	Score 314.6;	DB 17;	Length 1269;
Best Local Similarly	57.68;	Pred. No. 7.1e-46;		
Matches 732; Conservative	0;	Mismatches 479;	Indels 60;	Gaps 7

[illegible]

Dd	611	cggtgagaaatcttcttaagagagcgccgaacgcctcccgctctcaacgcgtgagatcagtgcaag	670
Qy	827	gcttcgttcaaccaacaacgagcgaattacaacgcctgttgaagagccgttcatatgacccgaacc	886
Dd	671	gcttcatcttcaacaacgagccaaactatctggtccacatgttgagacgttacccttgagacgaag	730
Qy	887	agcaagttgcggagacagccggtgtgagatgcggcggaatttctaacaggttgagatctctgacatg	946
Dd	731	gcacggtttaaaccgagccagactcatccgcgcagatcccaagtgtggtttgacttggaaaccagatccg	790
Qy	947	acgaagaccgaactacgacgtgttgactgttactctgagcttcgttcgcgtcgatgttttccaagca	1006
Dd	791	acgagctctctcttcgttccagagacctgctgtagcctctgattcgcgaagggcttccgttcgc	850
Qy	1007	gcatctgcgatgtctatctgaaacactttagaagcaacgctgttgagggtgtgtccgaaacgaacaa	1066
Dd	851	acatctgagatgtctatctgaaacactcccgcaacggtctgggggttggtccgaacccgttccagccg	910
Qy	1067	gcccagagacacgacgacgcatgttcaaacactctgtcaacacagctcgaagtgaattgaccttgcgc	1126
Dd	911	gaccgagctctcttccacccgacacttcaacacctacgtttagcagagagccgattaccgcgcgtta	970
Qy	1127	agcaacgcgcgcctgtgttgtaaacaccagaacggtgcggggccttcggccgaaccgcgcagagcaa	1186
Dd	971	tccaccccggttaaatgtgtgcaaacacagccgctgtgcgggcctgcgcgcgcgcgcgcgcgcgc	1022
Qy	1187	gcccagagacgaattctccgaacagcgcaacctgcgcgcgtatgtcttgatgaacagccgcgcgggtg	1246
Dd	1024	-----aaggttaaccgcgcctcccggtgttgtagcgcctacgtctgggtgagaaagcccgcgggtg	1078
Qy	1247	agtcggaacgagcaacagcgctgtgcgagcgatccggaacactgtgcgaagaagtgcgaaccccatgt	1306
Dd	1079	agtcggaacgagcgacgagagagatctccggaacagcaagagcaagggcttgagaccgcatgt	1138
Qy	1307	gcgcacgcgaactgaacagcagctgtgtaggggttactgacaaac-----ggcttaccga	1352
Dd	1139	gcgacccgacacttaccagaggaacgcgcgcgcgaacagcgcaacacccctcggtgtgcgtgcgcga	1196
Qy	1358	actcccgcatctgcgcgcgcagttggttcccgcgcgagtttagcaagctgtgtcgcgaacgac	1417
Dd	1199	agcgcgccatctcgcgcgcactggttctctgtccagttccgcagttccgcgcgcgcgcgcgcgcgc	1255
Qy	1418	ggcgaacgcggtg 1428	
Dd	1259	accgcgcctctg 1269	
RESULT 6			
AAAT94197			
ID	AAAT94197 standard; DNA; 3033 BP.		
AC	AAAT94197:		
XX			
DT	21-MAY-1998 (first entry)		
XX			
DE	Teredinibacter endoglucanase DNA.		
XX			
KW	Endoglucanase; cellulase; carboxymethylcellulose; cellulose;		
KW	biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;		
KW	thermostable enzyme; thermophillic; glycosidase; ss.		
XX			
OS	Teredinibacter sp. (clone 426P1).		
XX			
PN	W09744361-A1.		
XX			
PD	27-NOV-1997.		
XX			
PF	22-MAY-1997; 97WO-US08793.		
XX			
PR	22-MAY-1996; 96US-0651572.		
XX			
PA	(RECO-) RECOMBINANT BIOCATALYSIS INC.		

PI Lam DE, Mathur EJ;
XX
XX WPI: 1998-018435/02.
DR P-PSDB; AAM34989.
XX
PT Endoglucanase(s), preferably from archaeal bacterium, AEPIT 1a -
PT useful to degrade carboxymethylcellulose and hydrolyse of
PT beta-1,4-glycosidic bonds in cellulose
XX
XX
PS Claim 3; Fig 1E; 164pp; English.
XX
CC This DNA sequence from *Teredinibacter* sp. (clone 42CP1) encodes an
CC endoglucanase (see AAM34989) that is able to degrade
CC carboxymethylcellulose and to hydrolyse the beta-1,4-glycosidic
CC bonds in cellulose, and which shows homology to the thermostable
CC endoglucanase (see AAM34985) of *archaeobacterium hydrothermal vent*
CC isolate AEPIT1a. The DNA can be used in the recombinant production
CC of the endoglucanase and as a probe to identify similar sequences.
CC 24 Endoglucanase polynucleotides (see AAT94193-216) are claimed.
CC These can be incorporated into plasmid or virus-derived vectors for
CC use in a claimed method of producing enzymes in transformed host
CC cells. The claimed endoglucanases (see AAM34985-W35008) can be used
CC to degrade cellulose for the conversion of plant biomass into fuels
CC and chemicals, for use in detergents, textiles, animal feed, waste
CC treatment, and in the fruit juice and brewing industries for the
CC clarification and extraction of juices.
CC
SQ Sequence 3033 BP; 701 A; 886 C; 763 G; 683 T; 0 other;

Query Match 6.8%; Score 251.2; DB 19; Length 3033;
Best Local Similarity 55.8%; Pred. No. 6.9e-35;
Matches 603; Conservative 0; Mismatches 438; Indels 39; Gaps 5;

QY 280 gtgtgttccacatattcgagagcgctgtgtagccgctcgtcgatcaacgagcgtc 339
DB 1876 gtatttgcaacgaacacctgtgtatgtagtgaatcgatcgagagcgct 1935
QY 340 aacggcgagccgctgtgagacatattcgagcgccgctcccgagcgagggagac 399
DB 1936 gctgagcggtatggctgtgagacacctggaagagcccttgaca-----a 1983
QY 400 acccttgaaagtcattagattgcatctacatcgtccggagcgagctgcggcgctc 459
DB 1984 gggcgagacctgtcatgtgtgtgtgtagacgtccaaacgctgactgtgcgactc 2043
QY 460 gctctcaagcggaactgcccgtacgagcgaggtgttgagacctatgaagcgagtac 519
DB 2044 gctctcaagcggtgaactgcgacatctcgaaagtggtctcaacatctacaagctcgactac 2103
QY 520 atcgatcgatctgagatctcctgagcaatcgaaagtaaccagagcgtcgatcgatcg 579
DB 2104 atcgacacctatcgttaaatcatcagcgaccccgatagcgaggtatcaaaatcgctgcg 2163
QY 580 atcattgagcgagctcgtcgtccaaacgcggtlaccaaatltgacatltcaagcgtgt-- 636
DB 2164 gttatcgaggtgagctcaccacgtcctaactcgttacaactltgagcgaaactgtgcg 2223
QY 637 ---gcaagcggtgtgcgtatctagagcaagcgatcgagtagcgctcagcaaatgtcac 693
DB 2224 gaagcaaatgtgtcgtcgagctacccgagcgatcgtlccatgcatctgactgtgagc 2283
QY 694 gcatctcgagaggtgtagcatctacatcgagcgagcgccatccggtgctgtgtgccc 753
DB 2284 aaaatcccaagtatattctctcgtggtatattgacacactcgaagctcgtggtctggaac 2343
QY 754 aataatgccaagcgatagctacaggaagtlccagaa---ggtlccatcagcgagcatcg 810
DB 2344 gacaaactcgagcaagcggttaacctgattatgaaagtgttgcaacccgtgttcgcgc 2403
QY 811 gtcaagagctcgagcgctcgtcaccacaacgagcaataacagcgctgtgaagagcg 870
DB 2404 attaaccacatcgcggttcgtctgactgaactcgcgtactacacgacctgtggaagacc 2463

QY 871 ttcatgac---cgccacccagcagctcgcgagcaagccggtggagtlcgcgaaattctac 927
DB 2464 ttcttcgacagcgcaacacctcgagcggtcgtgtagcagccgctcgtcttcgattctat 2523
QY 928 cagttgaatcttgcacatcgacgaagccgactacggtgtgactgttactcgtcggtctgc 987
DB 2524 gaattgagacagctacatcgagagaagaaaccttcgtgacagcttggcgttctgcatalgtac 2583
QY 988 ggcgtgtgcttccaaagcagctgcatctcattacgacaccttaccgaaggttgggt 1047
DB 2584 tcgaaggtatgcgaagctccatcgatagctgcatcagcagtaacgagctgggt 2643
QY 1048 ggtcgaagcaacgaagcgccgagcagcgagcagctgttcaaaccttcgtcaacag 1107
DB 2644 ggcctcgtgaggttcacatctgtagcagctcactccacaacacttgaaaccttcgttaagaa 2703
QY 1108 tcgaagatgtaccttcgtgacacgagcggtcgtgtgtgcaacgaagcggtgcggcctc 1167
DB 2704 tcagttatcgacgctcgtgagcagcgcggaactgtgtgtaacgagccttggtgtgcgc 2763
QY 1168 ggcagcgccgagcgaagcagcagcgagcttcccgagcgcaccccgagcggtatgtc 1227
DB 2764 tacgttcaacgctgacactctccaggtat-----galtgcttactgt 2808
QY 1228 tggatcaagcgcggtgtagtcgagcgacaaagcgctgcgagcgatccgacactgyc 1287
DB 2809 tgggtgaaacacaaaggtgtgctgtagcgtgttcgcatcttaactcgtgatcgtacct 2868
QY 1288 aagaagtcgaaccccatgtgacacccgacgtlaacagacgtcgtacgggtactgaaccaac 1347
DB 2869 aacgacccggaacaaacagcagcagcacaatgtgtgatcgttcgcagcaacactgtccaac 2928

RESULT 7
AAX55661
ID AAX55661 standard; DNA: 11706 BP.
XX
XX AAX55661;
AC
XX
DT 30-JUL-1999 (first entry)
XX
DE DNA sequence encoding truncated cellulases Cel B4/5 and Cel B5.
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1/2;
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing; ss.
OS Unidentified.
XX
PN EP921188-A2.
PD 09-JUN-1999.
XX
PE 15-SEP-1998; 98EP-0810919.
XX
PR 19-SEP-1997; 97US-0932571.
XX
PA (CLRN) CLARIANT FINANCE BVI LTD.
PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
PI Gabb MD, Morgan H, Williams DP;
XX WPI: 1999-315403/27.
XX P-PSDB; AAT13492.
XX
XX New truncated cellulase proteins, useful in detergents and for
XX producing 'stonewashed' denim
XX
XX Disclosure; Page 20-23; 65pp; English.
XX
XX The invention relates to a recombinant cellulase active protein free of
XX proteins of native thermophilic and alkaliphilic origin, comprising


```
QY 1559 cgcgcagccgcagctcgcagccgcagctcgcagctcgcagctcgcagccgc 1618
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3835 caccgacacccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 3894
QY 1619 ctccgctcgcagccgcagctcgcagccgcagctcgcagctcgcagctcgcag 1678
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3895 caccgacacccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 3954
QY 1679 gccgcctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgc 1738
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3955 cgcgcagccgcagctcgcagccgcagccgcagccgcagccgcagccgcagcc 4014
QY 1739 cgcgcagccgcagctcgcagccgcagccgcagccgcagccgcagccgcagcc 1795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4015 caggagacccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 4074
QY 1796 cggctgataacccgcagccgcagccgcagccgcagccgcagccgcagccgcag 1855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4075 cgcgcagccgcagctcgcagccgcagccgcagccgcagccgcagccgcagcc 4134
QY 1856 attctgcagccgcagctcgcagccgcagccgcagccgcagccgcagccgcag 1915
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4135 attctgcagccgcagctcgcagccgcagccgcagccgcagccgcagccgcag 4191
QY 1916 tgcacaactgcagccgcagccgcagccgcagccgcagccgcagccgcagccgc 1975
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4192 gtcgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 4251
QY 1976 tgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcag 2033
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4252 tgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcag 4311
QY 2034 ----gttcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgc 2089
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4312 ggcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgc 4371
QY 2090 gtcgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 2149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4372 ggcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgc 4431
QY 2150 ggcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgc 2206
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4432 agcatcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcag 4491
QY 2207 ggcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcag 2266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4492 ggcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcag 4551
QY 2267 atcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagccgc 2291
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4552 caccgacacccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 4576

RESULT 9
AAQ13001
ID AAQ13001 standard; DNA; 2977 BP.
XX
AC AAQ13001;
XX
DT 14-OCT-1991 (first entry)
XX
DE Endol gene encoding endoglucanase.
XX
KM Cellulase activity; detergent; ds.
XX
OS Bacillus spp. NCIMB 40250.
XX
FH Key Location/Qualifiers
FT CDS 677..2779
FT sig_peptide 677..769
FT /tag= a
FT /tag= b
```

```
FT mat_peptide 770..2776
PN WO9110732-A.
XX
PD 25-JUL-1991.
XX
PF 18-JAN-1991; 91MO-DK00013.
XX
PR 19-JAN-1990; 90DK-0000164.
XX
PA (NOVO ) NOVO NORDISK A/S.
XX
PI Jorgensen PL, Schuelein M, Hansen C;
XX
DR WPI, 1991-238020/32.
DR P-PSDB; AAR13227.
XX
PT Enzyme exhibiting cellulase activity from Bacillus sp. - is an
PT endo-glucanase, esp. useful for harshness redn. of cotton-contg.
PT fabrics.
XX
XX Claim 14; Page 81; 96pp; English.
XX
CC The DNA (SEQ ID NO: 1) was sequenced from plasmid pPL517 which was
CC prep'd. from genomic DNA isolated from Bacillus spp. PL236. The
CC sequence contains three potential initiation sites at positions
CC 677, 737 and 749, but only the ATG at 677 is preceded by a ribo-
CC some binding site. The DNA may be inserted into an expression
CC vector for prodn. of recombinant endoglucanase in E. coli cells.
CC The enzyme is useful as a cellulytic agent.
CC See also AAQ13002 and AAQ13003.
CC
XX
SQ Sequence 2977 BP; 775 A; 737 C; 821 G; 644 T; 0 other;

Query Match 3.2%; Score 116.4; DB 12; Length 2977;
Best Local Similarity 57.3%; Pred. No. 1,6e-11;
Matches 251; Conservative 0; Mismatches 181; Indels 6; Gaps 2;

QY 1769 tgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgc 1828
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 148 tgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 207
QY 1829 tgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 1888
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 208 tcaaaacacgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 267
QY 1889 cccgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcag 1948
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 268 ccaagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 324
QY 1949 ggcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 2008
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 325 ggcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 384
QY 2009 tgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 2068
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 385 tgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 444
QY 2069 accgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcag 2125
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 445 tgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 504
QY 2126 ggcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 2185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 505 ggcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 564
QY 2186 tgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 2203
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 565 tgcagctcgcagccgcagccgcagccgcagccgcagccgcagccgcagccgcagcc 582

RESULT 10
```


[illegible][illegible]

RESULT	15
AA	AA546648/c
ID	AA546648 standard; DNA; 6510 BP.
XX	
AC	
XX	AA546648;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Tumour suppressor gene derived chemically modified sequence #370.
XX	
KW	Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW	cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW	cytosine methylation; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200168912-A2.
XX	
PD	20-SEP-2001.
XX	
PF	15-MAR-2001; 2001WO-EP02955.
XX	
PR	15-MAR-2000; 2000DE-1013847.
PR	06-APR-2000; 2000DE-1019058.
PR	07-APR-2000; 2000DE-1019173.
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI; 2001-602752/68.
XX	
PT	Fragments of chemically modified genes associated with tumour suppressor
PT	genes and oncogenes, useful in designing primers and probes for
PT	analysing diseases associated with cytosine methylation state e.g.
PT	cancer
XX	
PS	Claim 1; SEQ ID No 370; 27pp; English.
XX	
CC	The invention relates to a nucleic acid comprising a sequence of 18
CC	bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC	bisulphite, of genes associated with tumour suppression and
CC	oncogenes having a sequence taken from 536 (actually 533 since
CC	numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC	(SS) and sequences complementary to (SS). The nucleic acid may be a
CC	peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2002, 08:58:18 ; Search time 2693.99 Seconds
(without alignments)
18471.959 Million cell updates/sec

Title: US-09-917-384-2
Perfect score: 3687
Sequence: 1 atgagcgcgaacccacacacatc.....tcagctgcacagccagctga 3687

Scoring table: IDENTITY_NUC
Gapop 10-0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estbta:*
2: em_estbhm:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hrc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	99.6	2.7	1056	12	CNS02URL
c 2	99.4	2.7	966	12	CNS02SPN
c 3	97.4	2.6	906	12	CNS01S6G
c 4	96.8	2.6	986	12	CNS01YTT
c 5	93.2	2.5	971	12	CNS04ARS
c 6	91.8	2.5	475	12	CNS04S6N
c 7	90.8	2.5	723	12	CNS01UT4
c 8	90.8	2.5	1061	12	CNS03Y86
c 9	89.2	2.4	899	12	CNS02FE2
c 10	86	2.3	880	12	CNS01VOE
c 11	85.4	2.3	803	12	CNS026N7
c 12	85.2	2.3	613	12	CNS02D1N
c 13	84.2	2.3	949	12	CNS031R8
c 14	84	2.3	559	12	BH306809
c 15	83	2.3	744	9	BH611425
c 16	82.6	2.2	460	12	CNS036OU
c 17	82.6	2.2	873	9	BE217030

c 18	82	2.2	614	12	CNS01WDO	AL170277 Tetraodon
c 19	81.6	2.2	890	12	CNS035KB	AL228835 Tetraodon
c 20	81.6	2.2	1101	12	CNS04LE7	AL296008 Tetraodon
c 21	81.2	2.2	955	12	CNS02Y44	AL219397 Tetraodon
c 22	79.2	2.1	965	12	CNS01ZDK	AL174161 Tetraodon
c 23	79	2.1	616	12	CNS03VYH	AL262934 Tetraodon
c 24	78.4	2.1	403	12	CNS04F5S	AL287929 Tetraodon
c 25	77.6	2.1	441	9	BE125089	BE125089 DGL16_C1
c 26	77.4	2.1	611	12	BH379958	BH379958 AG-ND-102
c 27	77.2	2.1	478	10	BE496065	BE496065 WHE1261_G
c 28	77.2	2.1	523	10	BE145280	BE145280 WHE1834_F
c 29	76.8	2.1	822	12	AG154144	AG154144 Pan tlog1
c 30	76.6	2.1	760	12	CNS01NAQ	AL151987 Anopheles
c 31	76.4	2.1	401	10	BE332633	BE332633 us25402_Y
c 32	76.2	2.1	745	12	CNS01N20	AL151673 Anopheles
c 33	75.8	2.1	640	10	BE427339	BE427339 PSR6267_I
c 34	75.8	2.1	925	12	CNS0091P	AL053013 Drosophila
c 35	75.6	2.1	461	9	AW922188	AW922188 DGL1346_F
c 36	75.6	2.1	935	12	CNS006XK	AL066051 Drosophila
c 37	74.2	2.0	665	10	B1722807	B1722807 1031064C0
c 38	74	2.0	823	9	AM448279	AM448279 BRY 1782
c 39	74	2.0	829	12	CNS035GN	AL228704 Tetraodon
c 40	73.4	2.0	625	10	BC103499	BC103499 RH122_37
c 41	72.8	2.0	710	12	CNS022S1	AL178570 Tetraodon
c 42	72.6	2.0	663	10	BF483682	BF483682 WHE2336_F
c 43	72.4	2.0	654	12	AZ937136	AZ937136 2M0195012
c 44	72.2	2.0	554	10	BC367845	BC367845 HVSME1001
c 45	72.2	2.0	1101	12	CNS017SY	AL108460 Drosophila

ALIGNMENTS

RESULT 1
CNS02URL/c 1056 bp DNA linear GSS 15-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 167F12 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION AL214842.1 GI:7873661
VERSION GSS: genome survey sequence.
KEYWORDS Tetraodon nigroviridis.
SOURCE Tetraodon nigroviridis.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 1056)
Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billaud,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

TITLE Unpublished
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1056)
Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

REFERENCE 3 (bases 1 to 1056)
JOURNAL Unpublished
AUTHORS Direct Submission
TITLE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
source location/Qualifiers
1..1056
/organism="Tetraodon nigroviridis"

```

/db_xref="taxon:39883"
/clone_1b="G"
/clone_1b="G"
/note="Genoscope sequence ID : C0AG167DC06LP1-end : T7
190 a 358 c 273 g 234 t 1 others
BASE COUNT
ORIGIN

```

Query Match	2.7%;	Score 99.6;	DB 12;	Length 1056;
Best Local Similarity	56.8%;	Pred. No. 0.00032;		
Matches 183; Conservative	0;	Mismatches 139;	Indels 0;	Gaps 0

by 1452 gccgcctccgacggcagatccgtcgcttcgccgaatccgagcccgaatccgagccccgag 1511
||| ||||| ||||| | | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1019 GCtGAGTCCGAGGCCCGGGGCACAGAGCCCGGGCTGAGCCTGAGGTCCGGGCTGAGCCTGAG 960

Db 959 TCTGAGGCTGAGTCCGAGGCGCGGCGCGAGGCTGAGTCCGAGGCCCGGGCTGAGGCTGAG 900

[illegible]

Db 839 GC1GAGGCTGAGTCTGAGGCTGAGTCCGAGGCCGGGGCCGAGGCTGAGTCCGAGGCTTAC 780
 1692 gccgagcccgagtcgcgaagcccgagtaqctgcgcgcgcgcgtccgacgltcgtccgcqat 1755

Db 779 GCTGAGGGCTGAGGCTGAGTCCGAGGCGTGAAGCTGAGTCGAGAGGCCGGGGCTGAG 720
Gy 1752 gtccggtggctgaaggtgcag 1773

```

DD      /19 GCGGGGACCGAGGCTGAGGCTG 698
RESULT  2

```

LOCUS DEFINITION	CNS02SPN	966 bp	DNA	linear	GSS 15-MAY
	Tetraodon nigroviridis genome survey sequence T7 end of clone				
	162H05 of library G from Tetraodon nigroviridis, genomic survey				

ACCESSION	AL212180
VERSION	AL212180.1
KEYWORDS	GI:7870999
SOURCE	GSS: genome survey sequence.
ORGANISM	<i>Tetradodon lineatus</i>

ORGANISM
Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Actinopterygii; Teleostei; Euteleostei; Neoteleostei
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes

REFERENCE
AUTHORS
1 (bases 1 to 966)
Roest-crolius, H., Jallion, O., Dasilva, C., Filames, C., Fisher, C.
Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and

TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>
JOURNAL	Unpublished
REFERENCE	2. (bases 1 to 966)

TITLE	AUTHORS
Human gene number estimate provided by genome wide analysis using	Saurin, W. and Weissbach, J.
	Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
	Roest-Ciolinus, H., Jallion, O., Dastiv, C., Bonneau, L., Fisher, C.,
	Roest-Ciolinus, H., Jallion, O., Dastiv, C., Bonneau, L., Fisher, C.,

JOURNAL Unpublished
REFERENCE 3 (bases 1 to 966)
AUTHORS Genoscope.

COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at

http://www.genoscope.cns.fr/Tetraodon.
FEATURES
 Location/Qualifiers
 1. 966
 source

```

/clone_11b="G"
/clone="162H05"
/ab_xref="taxon:93883"
/Note="Genoscope sequence ID : C0AG162CD031r1-end : T7"

```

[illegible]

	150	300	450	600	750	900	1050	1200	1350	1500
Best Local Similarity	52.0%	Pred.	No.	0.00034						
Matches 223; Conservative	0;	Mismatches	206;	Indels	0;	Gaps	0			

Db	51	CGGACTCTTCTGATGCTCCGAGGCTAGGCCGCGAGACCAGGAGGCTGAGGCTGAGCTCCGAGGGCCG	110
QY	1510	agcagctccgccatccacccctccacccctctcccaagctcgaagcccatctcccatcccaaacacccca	1569

Db	111	GGGCGAGAGCGCGGGACCGAGCGCTGAGCGTGAATCCGAGCGCCGGGCGTGAAGCTGATGCCG	170
Qy	1570	agtcgagcccgagtagctgcgcgctgcgcgctctccgagctcgagcccgctctcgcctgcgcg	1629

D _b	171	AGTCCGAGCGCCGGGGCTGAGGCCTGAGGCCGGGGGCTGAGCGCTGACATCCGAGAGCCCGGGCAG	230
Q _y	1630	agcccgagtcctcgaagcccagatgactgcgcgcgttcgcagtctcgaagatcgaagccgcgtcgcg	1689

LZ	201	ACGCGGGAGCCGAGGC	TCGATC	TGATAT	CCGAAGGCCGGGGT	GTAAATC	GATATCCGAGAGCCG	290
QY	1690	TCGCGAGACCGAGT	CCGAGG	CCCGAGTAGCT	CAGCCGTC	CGCTCTCGAAGCT	CGTCGCGC	1749

[illegible]

Ox 1810 atcaaccgggtctccagttgtgaataaccgggtcgtcgtcgtgtgattgtcgaagt 1869
Db 411 AGGCTGAGCGCTGAGGCTGAGGCTGAGGCGGAGCCGAGGCTGAGGCGGAGCTGAGGCGG 470

QY	1870	acggtgcg	1878
Db	471	AGGCTGAGG	479

RESULT	3
CNS01S6C/G	
LOCUS	
CNS01SEC	
906 bp	
DNA	
148027	
DEC 13, 2000	

DEFINITION	tetracodon nigriviridis genome survey sequence PUC-ORI end of clone 169C18 of library G from <i>Tetracodon nigriviridis</i> , genomic survey sequence.
ACCESSION	AL164833

KEYWORDS	GSS; genome, survey sequence.
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis

REFERENCE
1 (bases 1 to 906)

TITLE Characterization and repeat analysis of the compact genome of the
Weissenbach, J.
Bonneau, L., Billault, A., Queller, F., Saurin, W., Bernot, A. and
Weissenbach, J.
Rochest, C., Collins, R., Vailionis, O., Lesiva, C., Zizanes, C., Fisher, C.,
Ardiansyah, N.

JOURNAL, REFERENCE AUTHORS
Unpublished
2 (bases 1 to 906)
Roest-Crollius, H., Jallou, O., Dasilve, C., Bouneau, L., Fisher, C.,

[illegible]

TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 986)					
AUTHORS	Roest-Crolius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzmes,C., Mincker,P., Brotlier,P., Quetier,F., Saunrh,W. and Weissenbach,J.					
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence					
JOURNAL	Unpublished					
REFERENCE	3 (bases 1 to 986)					
AUTHORS	Genoscope.					
COMMENT	Direct Submission Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases This sequence is a single read and was generated as part of a large scale clone-and sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon .					
FEATURES	location/Qualifiers					
SOURCE	1..986					
	/organism="Tetraodon nigroviridis"					
	/db_xref="taxon:99883"					
	/clone="219I03"					
	/clone.lib="G"					
	/note="Genoscoope sequence ID : C0AG219AE02LP1-end : T7"					
BASE COUNT	177 a 335 c 250 g 220 t 4 others					
ORIGIN						
Query Match	Best Local Similarity 2.6%; Score 96.8; DB 12; Length 986; Matches 191; Conservative 54.9%; Pred. No. 0.00074; Mismatches 157; Indels 0; Gaps 0;					
OY	1452	gccgcgtccgcccgcagatccgctcgcgagtcgcgcgagcccgagcccgagcccgag	1511			
Db	950	gctgagtccgaaggccggaccgacgactgagctgagctgagctgagccgacgagctgag	891			
OY	1512	cagctcgccatcgccgctcgccgctcgcgagtcgcgagcccgctcgcgagcccgag	1571			
Db	890	tctgagtgctgagagccgggctgagtcgagccgagccgggctgagctgagctgagctgag	831			
OY	1572	tccgagcccgagtagtgcgcgctcgccgctcgcgagtcgcgagcccgctcgcgagcccgag	1631			
Db	830	tccgagcccgagccgagtcgagccgagccgagccgagccgagccgagccgagccgag	771			
OY	1632	ccccgagtcgcgagcccgagtagctgcgcgctcgccgctcgcgagtcgcgagcccgctcgc	1691			
Db	770	gctgagtgctgagagccgggctgagcccgagccgagccgagccgagccgagccgag	711			
OY	1692	gccgagcccgagtcgcgagcccgagtagctgcgcgctcgccgctcgcgagtcgcgagcccgctcgc	1751			
Db	710	gctgagtgctgagagccgggctgagccgagtcgagctgagctgagctgagccgagccgagccgag	651			
OY	1752	gtcgagtgctgagtagtgcagatacaagaacaagatccgagcccgag	1799			
Db	650	gccggagccgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag	603			
RESULT	5					
LOCUS	CNS044R5	971 bp		DNA	linear	GSS 18-MAY-2000
DEFINITION	Tetraodon nigroviridis genome survey sequence PUC-Orf end of clone 08AG23 of library G from Tetraodon nigroviridis, genomic survey sequence.					
ACCESSION	AL274442					
VERSION	AL274442.1 GI:7996717					
KEYWORDS	GSS: genome survey sequence.					
SOURCE	Tetraodon nigroviridis. Tetraodon nigroviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciforma; Tetraodontiformes; Tetraodontidae; Tetraodon.					

[illegible]

ORGANISM	Tetraodon nigroviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorphia; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE	1 (bases 1 to 475) Roest-Crollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
AUTHORS	Unpublished
JOURNAL	2 (bases 1 to 475) Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brotier,P., Queller,F., Saurin,W. and Weissenbach,J. Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
TITLE	Unpublished
JOURNAL	3 (bases 1 to 475) Genoscope.
REFERENCE	Direct Submission
AUTHORS	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .
JOURNAL	Location/Qualifiers
COMMENT	1..475 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone="005024" /clone_1lb="H" /note="Genoscoope sequence ID : COBH005DE12XD1-end : T7"
FEATURES	source
BASE COUNT	70 a 116 c 227 g 55 t 7 others
ORIGIN	
Query Match	2.5%, Score 91.8; DB 12; Length 475;
Best Local Similarity	56.9%; Pred. No. 0.0034;
Matches 182; Conservative 2; Mismatches 135; Indels 1; Gaps 1;	
OY 1453	ccgcctccgccgcgcagatcgcgttcgcgtcccgagatccgaaccgagatccgacccggagc 1512
Db 153	CCGAGGCTCGAAGCCTGAGCTGAGGCTGAAGGCCGGGGCCGAGGCTGTATCCGAGCGGGG 212
OY 1513	agctgcgatccgcgtccgcgtctccgagatccgagcccgctccc-gtgcgcgagccggag 1571
Db 213	CTGAGGCCGGGGGCTTAGCTGATTCCGAGGCCGGGGGCTGAGGCCGCGTGCTGAAGCTAG 272
OY 1572	tccgagcccgagtagctgcgcgtccgcgtctccgagatccgagcccgctctccgtccgcgag 1631
Db 273	TCCGAGGCCGGGGGCTGAAGGCCGGGGCCGAKCGTGAGTCGAGGCCGGGGGCTGAAGGCCGG 332
OY 1632	cccgagatccgagcccgagtagctgcgcgtccgcgtctccgagatccgagcccgctccgctc 1691
Db 333	ACCGAGGCTGAGGCCGAGAGCTKAGGCCCGGAGACCAGGCTGAGGCTKAGGCTTAGCTAG 392
OY 1692	gccgagcccgaggttcgcgaaccgagtagctgcgcgtccgcgtctccgagatcgttcgcggt 1751
Db 393	GCCGAGGCTGAGTCCKAAGCCGGGGCTTAGAGCCGGGAGCCGAGGCTGAAGGCCGAGGCTGAT 452
OY 1752	gtcggagtggactgaaggtgc 1771
Db 453	GCCGGGACCGAGGCTGANGC 472
RESULT 7	CNSOJUT4 723 bp DNA linear CSS 12-MAY-2000
CNSOJUT4/c	Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
LOCUS	197624 of library G from Tetraodon nigroviridis, genomic survey
DEFINITION	sequence.

ACCESSION	AL168241.1
VERSION	GI:7806298
KEYWORDS	GSS: genome survey sequence.
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodon. 1 (bases 1 to 723) Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C., Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis Unpublished 2 (bases 1 to 723) Roest-Crolius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J. Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence Unpublished 3 (bases 1 to 723) Genoscope. Direct Submission Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases This sequence is a single read and was generated as part of a large scale clone-and sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon . Location/Qualifiers 1..723 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone="197G24" /clone_1lb="c" PUC-ori"
FEATURES	Source
BASE COUNT	109 a 330 c 170 g 110 t 4 others
ORIGIN	
Query Match	2.5%, Score 90.8; DB 12; Length 723;
Best Local Similarity	53.7%; Pred. No. 0.0046;
Matches 188; Conservative	0; Mismatches 162; Indels 0; Gaps 0;
Oy 1424	cgggtcgcaagctgcgacccgcttcgagcccgccgctccgcgcgcgcgagtccgtcgttcgc 1483
Dd 436	CGGGCTAGGCTGAGTCGTGTCYGGGCCGGGGCTGAGGCTAGGCTGAGGCTGAAGCCGGGGC 437
Oy 1484	cgagtcgagcccgagtcgagcccgagccgagcagctgcgcatcgccgctcgccgttcaggct 1543
Dd 436	TGAGCTTAGATCCGGGGCTGAGGCCGAGGCTGAGGCCGAGGCTGAGGCCGGGGCTGAGGC 377
Oy 1544	cgagccgcttctcgttcgcgcgagcccgagtcgcgagcccgagtagatcctcgcgcttc 1603
Dd 376	CGAGGCTGAGGCCGGGGCTGAGGCTGAAGCTGAATCCGGGGCTGAGGCCGGGGCTGAGGC 317
Oy 1604	cgagtcgagccgcttcgttcgcgcgagcccgagtcgagcccgagtagatcgtccgctgc 1663
Dd 316	CGAGGCTAGGCCGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGC 257
Oy 1664	cgctcgcagtcgagcccgcttcgctgcgcgagcccgagtcgagcccgagtagactgc 1723
Dd 256	TGAGGCCGGGACGTGATGTGAGTCCGGGGCTGAGGCCGAGGCTGAGGCCGAGGCTGAGGC 197
Oy 1724	cgctcgcgcttcgcgagtcgctgcgcgagtcgagtgagtgagtgag 1773
Dd 196	CGGGCTAGGCTGAGGCTGAAGGCTGAATCCGGGGCTGAGGCTGAGGCTG 147

[illegible]

Db 664 CGGGGCTGAGGCTGAGGCTGAGTCCGGGGCTGAGGCTGAGGCTG 615

RESULT 9
CNS02F22/c
LOCUS
DEFINITION
CNS02F22 899 bp DNA linear GSS 13-MAY-2000
Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
11311 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL194483.1 GI:7832589
GSS: genome survey sequence.
Tetraodon nigroviridis.
Tetraodon nigroviridis.
Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 899)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 899)
Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizames,C., Winkler,P., Brotlier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 899)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.
Location/Qualifiers
source
1. .899
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="11311"
/clone_1lb="G"
/note="Genoscope sequence ID : C0AG131CE06SP1-end :
PUC-ori"

BASE COUNT 145 a 340 c 231 g 180 t 3 others

ORIGIN

Query Match 2.4%; Score 89.2; DB 12; Length 899;
Best Local Similarity 52.7%; Pred. No. 0.0074;
Matches 193; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

Db 1408 gcgagcgacgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1467

Db 828 GAGGCCGAGCCGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 769

Db 1468 agtcgctgagctcgagcgagcgagcgagcgagcgagcgagcgagcgag 1527

Db 768 AGGCCGGGGCTGAGGCTGAGTCCGAGGCCGGGGCTGAGAGCCGGGGCTGAGGCTG 709

Db 1528 tcgagcgctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1587

Db 708 AGGCTGAGTCTGAGGCTGAGTCCGAGGCCGGGGCTGAGGCTGAGTCTGAGTCCG 649

Db 1588 tcgagcgctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1647

Db 648 AGGAGGGGGCTGAGAGCCGGAGCCGAGGCTGAGAGCCGGAGCCGAGGCTGAGGCTG 589

Db 1648 agtagctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1707

Db 568 AGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 529

Db 1708 agccgagtagctcgagcgagcgagcgagcgagcgagcgagcgagcgag 1767

Db 528 AGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 469

Db 1768 gtcgag 1773

Db 468 GGGCTG 463

RESULT 10
CNS01VOE/c
LOCUS
DEFINITION
CNS01VOE 880 bp DNA linear GSS 12-MAY-2000
Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
199E01 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL169367.1 GI:7807424
GSS: genome survey sequence.
Tetraodon nigroviridis.
Tetraodon nigroviridis.
Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 880)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 880)
Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizames,C., Winkler,P., Brotlier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 880)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.
Location/Qualifiers
source
1. .880
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="199E01"
/clone_1lb="G"
/note="Genoscope sequence ID : C0AG199AC01SP1-end :
PUC-ori"

BASE COUNT 160 a 324 c 223 g 172 t 1 others

ORIGIN

Query Match 2.3%; Score 86; DB 12; Length 880;
Best Local Similarity 53.6%; Pred. No. 0.02;
Matches 179; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

Db 1424 cgtgctgagctcgagcgagcgagcgagcgagcgagcgagcgagcgag 1483

Db 832 CGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 773

Db 1484 cgagtcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1543

Db 772 CGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 713

Oy	1544	cgaagccgcttcgcgtcgacggagcccgaaatcgcagaccgafragtccgcgcgtccgtc	1603
Dd	712	TGAGGCTTAGGGCTGAGGCTGAGGCCGGGGCGTGA GTCTATGTCCGGGGCGGAGTCTGAGTC	653
Oy	1604	cgaactcgaagccgcgtcttcgcgtcgacggagcccgagctcgaagccagaatagctcgccgtccg	1663
Dd	652	CGGGGCTGAGGGCCCAAGGCTGAGACCCGGGGCTTGAGGCTTAGAGCTGAGGCTGAGGCCGGGGC	593
Oy	1664	cgctccgagctcgaagcccgcttcgcgtcgacggagcccgagctcgaagcccgagtagctcgc	1723
Dd	592	TGAGCTGATGATCGGGGCTGAGTCTGAGTCTCGGGGCTGAGGCTGAGGCTGAGGC	533
Oy	1724	cgctgcgcttcgcgaagctcgtcgcgcggtgtcggg	1757
Dd	532	CGAGGCTGAGGCCCGAGGCTGAGGCCGGGGCTGAG	499
<hr/>			
RESULT 11			
<hr/>			
CNMS026N7			
LOCUS			
DEFINITION			
Tetraodon nigriviridis genome survey sequence PUC-Orl end of clone			
241C10 of library G from Tetraodon nigriviridis, genomic survey			
sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
<hr/>			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
COMMENT			
<hr/>			
FEATURES			
source			
1..803			
/organism="Tetraodon nigriviridis"			
/db_xref="taxon:99883"			
/clone="241C10"			
/clone_1lb="G-			
/note="Genoscope sequence ID : CGAG241B05SP1-end :			
PUC-Orl"			
Location/Qualifiers			
BASE COUNT			
ORIGIN			
95 a 218 c 223 g 192 t 75 others			
<hr/>			
Query Match			
Best Local Similarity 52.1%; Pred.No.0.024;			
Matches 139; Conservative 18; Mismatches 110; Indels 0; Gaps 0;			
<hr/>			
Oy	1449	ccccgcgcctccgccgcgcgagctcgttgcgttcgccgagctcgaagcccgagctcgaagccc	1508
Dd	235	CACGGCGCTCCACGNCNTAGCAGCGCTCGCACAGDAGGCCCTCGA VCTGNSGCCGACCCT	294

Oy	1509	gagcgcacgcgcacatcgccgttcgcgcctccgcagcctcgagccgcgtctccgttcgcagaccg	1568
Db	295	GTSCCTGTSCTGTATVCTCGRRCCCGACCTTAASCTGTARCCCGAAGCTGTACTGTSTCT	354
Oy	1569	gagtcgcagcccgagtagctgcgcgttcgcgcgtctccgcagcctcgagaccgcctctccgtgcc	1628
Db	355	GGGCCCCGACCCTGTGTCCTGCTGCTGCCTGAGGCCCGGGGCGCCGAGCCGAGCCTGTGCTGAGCCC	414
Oy	1629	gagccgcagatccgcagaccgcagtagctgcgcgttcgcgcgtctccgcagctcgagcctcgacctcc	1688
Db	415	GASCCCGNNSCCCGGSGCCTGAGACCTGTGCTGTGCTGATGCTCGAACCTCGRRCGCCGACCTT	474
Oy	1689	gtcgscgcagcccgagtagtccgagccccgag	1715
Db	475	GTSCCGGGGCGCCGNMNCCTGAGCCTGTGTG	501
RESULT 12			
CNSOZD1N	613 bp DNA linear GSS 13-MAY-2000		
LOCUS	Tetraodon nigroviridis genome survey sequence T7 end of clone		
DEFINITION	256B21 of library G from Tetraodon nigroviridis, genomic survey		
sequence.			
ACCESSION	AF191876		
VERSION	AF191876.1 GI:7829980		
KEYWORDS	GSS; genome survey; sequence.		
SOURCE	Tetraodon nigroviridis.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.		
REFERENCE	1 (bases 1 to 613)		
AUTHORS	Roest-Criollius,H., Jallion,O., Dasilva,C., Filames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.		
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 613)		
AUTHORS	Roest-Criollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Filames,C., Wincker,P., Brotlier,P., Quetier,F., Saurin,W. and Weissenbach,J.		
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 613)		
AUTHORS	Genoscope.		
JOURNAL	Direct Submission		
TITLE	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases		
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon.		
FEATURES	Location/Qualifiers		
SOURCE	1..613		
	/organism="Tetraodon nigroviridis"		
	/db_xref="taxon:99883"		
	/clone_"256B21"		
	/clone_"11b-G"		
	/note="Genoscope sequence ID : COAG256CA11bP1-end : T7"		
BASE COUNT	114 a 123 c 264 g 100 t 12 others		
ORIGIN			
Query Match	2.3%; Score 85.2; DB 12; Length 613;		
Best Local Similarity	54.9%; Pred. No. 0.025;		
Matches 168; Conservative	0; Mismatches 138; Indels 0; Gaps 0;		
Oy	1452	gcgcgcctcgcgcgcgcagatcgcgttcgcgcgttcgcgcagatccgagccgagtcgcagccgag	1511
Db	177	GGCGGGCGGTGAGGCGTGAGTCCGGGGCTGAGGCTGAGTCCGGGGCTGAGGCTGAGGCTGAG	236

QY 1512 cagctcgccatcgccgttcgcttcgcagctcgagcccgcttcgcttcgcagcccgag 1571
DB 237 GCTGAGCGCGGGGCTGAGGCTGAGTCGCCGGGCTGAGGCTGAGGCTGAGGCTGAGGCGGG 296
QY 1572 tccgagcccgagtagctcgccgcttcctcgcagctcgagcccgcttcgcgcgcgag 1631
DB 297 GCTGAGGCTGAGTCCGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 356
QY 1632 cccgagctccgagcccgagtagctcgccgcttcctcgcagctcgagcccgcttcgcgc 1691
DB 357 TCCGGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 416
QY 1692 gccgagcccgagctccgagcccgagtagctcgccgcttcgcagcttcgcagcttcgcgcg 1751
DB 417 GCTGAGGCGGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 476
QY 1752 gtcgag 1757
DB 477 GCCGGG 482

RESULT 13
CNS031R8 949 bp DNA linear GSS 15-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 205022 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL223901.1 GI:7882739
VERSION GSS; genome survey sequence.
KEYWORDS Tetraodon nigroviridis.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 949)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 949)
Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fitzames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 949)
Genoscope.
REFERENCE Direct Submission
AUTHORS Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
JOURNAL This sequence is a single read and was generated as part of a large
COMMENT scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source 1..949
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="205022"
/clone_1lb="G"
/note="Genoscope sequence ID : C0AG205BH11P1-end : T7"
BASE COUNT 205 a 202 c 390 g 144 t 8 others
ORIGIN
Query Match 2.3%; Score 84.2; DB 12; Length 949;
Best Local Similarity 49.0%; Pred. No. 0.034;
Matches 221; Conservative 1; Mismatches 229; Indels 0; Gaps 0;

QY 1428 gccagcttcgacagctcgagcccgcccttcgcgcgcgagctcgcttcgcgcgcgag 1487
DB 371 GCCGAGGCTGAGACCTGAACTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 430
QY 1488 tccgagcccgagcttcgagcccgagtagctcgccgcttcgcgcgcgagctcgag 1547
DB 431 CCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 490
QY 1548 cccgcttccttcgagcccgagcccgagtagctcgccgcttcgcgcgcgagcttcgcgc 1607
DB 491 ACCGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 550
QY 1608 ctcgagcccgcttcgcttcgcgcgcgagcccgagcttcgcgcgcgagcttcgcgcgc 1667
DB 551 GTCGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 610
QY 1668 tccgagcttcgagcccgcttcgcgcgcgagcccgagcttcgcgcgcgagcttcgcgcgc 1727
DB 611 ACCGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 670
QY 1728 gccgcttcgagcttcgagcccgagcttcgcgcgcgagcttcgcgcgcgagcttcgcgcgc 1787
DB 671 GCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 730
QY 1788 ttcgagcccgagcttcgagcccgagcttcgcgcgcgagcttcgcgcgcgagcttcgcgcgc 1847
DB 731 GCCGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 790
QY 1848 gtcggtgattgttcgagccgagcttcgcgcgcgagcttcgcgcgcgagcttcgcgcgc 1878
DB 791 GACGGGACCSAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 821

RESULT 14
BH306809 559 bp DNA linear GSS 30-NOV-2001
LOCUS CH230-100P7, TY CHORI-230 Segment 1 Rattus norvegicus genomic clone
DEFINITION BH306809
ACCESSION BH306809.1 GI:17219217
VERSION GSS.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 559)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
A., Gebregiorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 ECORI segment
Unpublished (1999)
Other GSS: CH230-100P7.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat30.htm). For BAC library
availability, please contact Pieter de Jong (pjejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or_eing_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Seq primer: 97
Class: BAC ends.
FEATURES
source 1..559
/organism="Rattus norvegicus"
Location/Qualifiers

[illegible]

genes. *Genome Res.* 10 (10), 1617-1630 (2000)
Wagl, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Matsubara, M., Yonekura, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
Hayashizaki, Y., Okazaki, T., Muramatsu, M., Inoue, T., Kira, A. and
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system -384-format
sequencing pipeline with 384 multicapillary sequencer. *Genome Res.*
10 (11), 1757-1771 (2000)
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA
encyclopaedia: real-time sequence clustering for construction of a
nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)
Yamanaka, T., Kiyosawa, H., Kondo, S., Saito, T., Shingawa, A., Aizawa
, K., Fukuda, S., Hara, A., Itoh, M., Kawai, T., Shibata, K., Aikawa, T.,
Ishii, Y. and Hayashizaki, Y.

Mapping of 19032 mouse cDNAs on mouse chromosomes. *J. Struct.*
Func. Genomics 2 pre, U72-U86 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

e mouse tissues.

FEATURES					
	source	Location/Qualifiers			
OY	1744	/organism="Mus musculus"			
		/strain="C57BL/6J"			
		/dp_xref="taxon:10990"			
		/clone="3002003A21"			
		/clone_1lb="RIKEN full-length enriched, 12 days embryo head"			
		/sex="mixed"			
		/tissue_type="head"			
		/dev_stage="12 days embryo"			
		/lab_host="SDR"			
		/note="Site.1: XhoI; Site.2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'] GAGGAGAAGATCTTCGAGTTAAATTAAATCCCCCCCCC 3']"			
BASE COUNT	139 a	247 c	281 g	77 t	
ORIGIN					
Query Match	2.3%	Score 83;	DB 9;	Length 744;	
Best Local Similarity	58.8%;	Pred. No.	0.049;		
Matches 143;	Conservative %	0;	Mismatches 100;	Indels 0;	Gaps 0;
OY	1483	ccgagtccgagcccgagtcgcgagcccgagcgcgtgccatcgcgcttcgcgagc	1542		
Db	176	ccgaccCGGACCCCGAGCGGAGCCCGGAGCCCGAGCCGGAGGCCGAGCCGAGCCGGGCG	235		
OY	1543	tggagccgcttcgctcgccgagcccgagtcggagcccgagtagctgcgctcgccgct	1602		
Db	236	cggacCGGAGCCCATCGAGGCCCGAGCCCGAGACC CGAGCCGAGCCGACCGCCGAGC	295		
OY	1603	ccgagtcgagccccgcttcgcgccgagcccgagtcggaagcccgagtgcgcgtcg	1662		
Db	296	CCCAcCGGAGCCGGAACCCGAACAAGAAGCCGGAGCCCGACCGGAGCCGCAACCCGAGC	355		
OY	1653	ccgcttcgagtcgagcccgcttcgctgcgcggagcccgagtcggaagcccgagtgcg	1722		
Db	356	CCTTACAGAGAGCCTGAGCCTGAGCCTGAGACCTGAGACCTGAGACTGAGAGGCAGC	415		
OY	1723	ccg 1725			

Tue Sep 3 13:41:59 2002

us-09-917-384-2.rst

Page 10

Db 416 CCG 418

Search completed: August 31, 2002, 10:40:44
job time: 6146 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2002, 09:04:38 ; Search time 77.71 Seconds
(without alignments)
11654.237 Million cell updates/sec

Title: US-09-917-384-2
Perfect score: 3687
Sequence: 1 atgagagcgaaccacaacac.....tcagctgcacagcagctga 3687

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*
3: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6A_COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	427	11.6	3004	1 US-08-276-213-6	Sequence 6, Appl
2	324	8.8	3404	1 US-08-265-429A-1	Sequence 1, Appl
3	324	8.8	3404	1 PCT-US95-09069-1	Sequence 1, Appl
4	314.6	8.5	1269	1 US-08-265-429A-4	Sequence 4, Appl
5	314.6	8.5	1269	5 PCT-US95-09069-4	Sequence 4, Appl
6	134.4	3.6	11707	4 US-09-136-574A-1	Sequence 1, Appl
7	116.4	3.2	2977	2 US-07-862-588B-1	Sequence 1, Appl
8	98.8	2.7	6416	4 US-09-136-574A-2	Sequence 2, Appl
9	95.2	2.6	1116	4 US-09-104-308-2	Sequence 2, Appl
10	95.2	2.6	1116	4 US-09-321-981-2	Sequence 2, Appl
11	94.8	2.4	390	4 US-09-197-649-7	Sequence 7, Appl
12	87.8	2.4	4403765	4 US-09-103-840A-2	Sequence 2, Appl
13	87.8	2.4	4403765	4 US-09-103-840A-2	Sequence 2, Appl
14	81.6	2.2	2029	4 US-09-136-574A-6	Sequence 4, Appl
15	81.2	2.2	4411529	4 US-09-103-981A-1	Sequence 4, Appl
16	80	2.2	1716	4 US-09-321-981-4	Sequence 4, Appl
17	77.6	2.1	2712	3 US-09-025-691-4	Sequence 4, Appl
18	77.4	2.1	985	1 US-09-056-556-182	Sequence 182, App
19	77.4	2.1	2293	1 US-08-604-913B-12	Sequence 12, Appl
20	77.4	2.1	2293	1 US-08-604-913B-12	Sequence 12, Appl
21	74.2	2.0	1931	2 US-09-130-114-2	Sequence 1, Appl
22	70.4	1.9	4411529	4 US-09-103-840A-1	Sequence 1, Appl
23	69.4	1.9	1280	4 US-09-060-756-4	Sequence 4, Appl
24	68.2	1.8	535	4 US-09-056-556-171	Sequence 171, App
25	67.4	1.8	897	4 US-09-434-288-6	Sequence 6, Appl
26	67.4	1.8	2668	2 US-08-461-175-11	Sequence 11, Appl
27	67.4	1.8	2668	3 US-09-031-606-11	Sequence 11, Appl

C 28	66.8	1.8	494	4 US-09-056-556-176	Sequence 176, App
29	65.8	1.8	1248	4 US-09-105-537-7	Sequence 7, Appl
C 30	65.8	1.8	13613	4 US-09-105-537-3	Sequence 3, Appl
31	65.4	1.8	2219	3 US-08-510-646B-17	Sequence 17, Appl
C 32	65.4	1.8	5970	3 US-09-320-878-21	Sequence 21, Appl
33	65.4	1.8	28958	1 US-08-238-261B-6	Sequence 6, Appl
34	65.4	1.8	28958	1 US-08-456-837-6	Sequence 6, Appl
35	65.4	1.8	28958	1 US-08-457-342-6	Sequence 6, Appl
36	65.4	1.8	28958	1 US-08-457-646A-6	Sequence 6, Appl
37	65.4	1.8	28958	1 US-08-458-076A-6	Sequence 6, Appl
38	65.4	1.8	28958	1 US-08-764-233A-4	Sequence 4, Appl
39	65.4	1.8	28958	1 US-08-457-335A-6	Sequence 6, Appl
40	65.4	1.8	28958	1 US-08-729-214-6	Sequence 6, Appl
41	65.4	1.8	28958	3 US-09-028-934-6	Sequence 6, Appl
42	65.4	1.8	49377	1 US-08-764-233A-1	Sequence 1, Appl
43	65.2	1.8	43280	2 US-08-804-227C-1	Sequence 1, Appl
44	64.8	1.8	15872	4 US-09-105-537-1	Sequence 1, Appl
45	64.8	1.8	44377	2 US-08-804-227C-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-08-276-213-6
Sequence 6, Application US/08276213
Patent No. 553655
GENERAL INFORMATION:
APPLICANT: Thomas, Steven
APPLICANT: Laymon, Robert
TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276.213
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Edna
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: NREL IR# 94-08
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)231-1000
TELEFAX: (303)231-1098
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-276-213-6
Query Match 11.6% Score 427; DB 1; Length 3004;
Best Local Similarity 99.8%; Pred. No. 1.6e-73;
Matches 438; Conservative 0; Mismatches 0; Indels 1; Gaps 1;


```
Db 1904 GACCGAGCTCTCCACCGACTTCACACTTACGTTAGACGAGACCGCTATCCGACCGCTGA 1963
Qy 1127 agcaccgagcgtgtgtgtaaccaagaacggtgcgggacctcggccagcgccagcaaa 1186
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1964 TCCACCCCGCTTACTGTGTCACACACGACGCGGTCCGGCTCGCGGAGC----- 2011
Qy 1187 gcccgagagacttccgaaacgagacacttcgagcgtatgtctggaatcaagccgagtg 1246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2012 GCGCCACGCTCAACCCGGCTCCCGGTGTGACGCTTACGCTACGCTGGTGAAGCCCCGGGTG 2071
Qy 1247 agtcgagcagcacaagcgtctgagagatcccgacaacttgcgaagaagtcggaacccatgt 1306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2072 ACTCCGACGCGCCGACGAGAGATCCGAAACGACGAGGCGGCTTCGACCGGATGT 2131
Qy 1307 gccacccgagctacacagcagctgtacggggtactgaaccaac-----ggctaacga 1357
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2132 GCGACCCGACCTACACAGGCGACACGCGCCGCAACGCAACACCCCTGGGTGCGCTGCCA 2191
Qy 1358 actcccgagatcgccgagcagtggttcccgagcagtttgacagactgttcgagagcgac 1417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2192 ACGCCCCACTTCCGCGCACTGCTGCTGCGCCAGTTCCGCGAGCTGCTGCCAAGCCT 2251
Qy 1418 ggcacgagcgtgcgagcgtgcagcagctgcagagcccgccgctccgcccgcgagtcgtcg 1477
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2252 ACCCGGCTGTGTAACCGGAGAGTAGGCAACGGCTGACAGCTTCACAGGAGMACTGATCAG 2311
Qy 1478 ctccgagagtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1527
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2312 CACCTCTTACGCGGACGAGCGCGCCGCTCCACTCCCGTGGGCGGCGCGC 2361
```

RESULT 3

```
PCT-US95-09069-1
: Sequence 1, Application PC/TUS9509069
```

GENERAL INFORMATION:

```
: APPLICANT: Wilson, David B.
: APPLICANT: Walker, Larry P.
: APPLICANT: Zhang, Sheng
: TITLE OF INVENTION: Thermostable Cellulase From A Thermomonospora Gene
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
: STREET: 1800 One Met Plaza
: CITY: Buffalo
: STATE: New York
: COUNTRY: United States
: ZIP: 14203-2391
```

COMPUTER READABLE FORM:

```
: MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
: OPERATING SYSTEM: IBM compatible
: OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
: SOFTWARE: Wordperfect for Windows 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/09069
```

FILING DATE:

```
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: U.S. Serial No. 08/265,429
```

ATTORNEY/AGENT INFORMATION:

```
: NAME: Nelson, M. Bud
: REGISTRATION NUMBER: 35,300
: REFERENCE/DOCKET NUMBER: 18617.0008
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (716)856-4000
: TELEFAX: 716-849-0349
```

INFORMATION FOR SEQ ID NO: 1:

```
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3404 nucleotides
: TYPE: nucleic acid
: STRANDEDNESS: double-stranded
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
```

```
: ORIGINAL SOURCE:
: ORGANISM: Thermomonospora fusca
: STRAIN: YX36
: CELL TYPE: bacterium
: PCT-US95-09069-1
```

```
Query Match 8.8%; Score 324; DB 5; Length 3404;
Best local Similarity 56.6%; Pred. No. 8.5e-54;
Matches 775; Conservative 0; Mismatches 535; Indels 60; Gaps 7;
```

```
Qy 200 tcgtcaacccgtactggcgcaagaagttacagagcgaagcggagacacaaatgcca 259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1010 TCGACAACCCGTTTCGAGGGGGCCCAAGCTTACGTGAACCCGGTCTGGCCCAAGGCCG 1069
Qy 260 ctctcgagagaaatgctgtctgttccacatatatgaagcgccgtctgtatgagcga 319
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1070 CCGCTGAGCCGGGGGCTTCGCGGTCCGCAACGATCCACCGCTGTGCTGGTGAACCGTA 1129
Qy 320 tcgctcgatcaacgagcgtcaacgagcggaccggcgttgacacacatctgagccgcc 379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1130 TCGGCGCATCGAGGGCAAGAGACGCCGCAACCGGCTCCATGAGCTGCGGACACCC 1189
Qy 380 t-----ctccagcagcaaggaaacacccctgaagtcattgagattgcatcag 430
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1190 TGGAGAGGCGCTCCGCGCACTCGGTGCGACCGCTGACCATTCACAGTGTGCTATCTACA 1249
Qy 431 atctcgagagagcagagatgagcgagctgcgtctccaaagcggaactgccgctacggcag 490
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1250 ACCTGGCCGGCGCGATGGCGCGCGGTGCGCTCCACAGGTGAGCTG-----GTCCTCG 1303
Qy 491 caggtttcgagacatagaaacgagatcacatccagattgcagatctccgaacat- 549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1304 ATGACACTCGACCGCTTACAAAGCGAAGTACATTCAGACCCGATCGCGAATCATGTGGGACT 1363
Qy 550 --ccgaagtactccagcctgcgagtcgtgaagatcatttggccgagactcgtgccaaacg 607
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1364 TCGCAGACTACGAGAACCTCGGATCGTCCCATATCGAGATGACTCCCTGCCCAACC 1423
Qy 608 cgttcacacataatgagattccaagcgtgtgaaagcggtgcgtta----- 653
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1424 TCGTCAACCAAGTGGGGGGAACCGCGCGACCGAGCTCTGCGCTTACATGACGAGAACG 1483
Qy 654 ----ttacgagcaagcagcagtaagcgtccaaagaattgcagccatccgaagtg 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1484 GCGGCTTACGTCACAGGTGTGGGCTACCGCCTCCGCAAGCTGGGGAATCCGAACTCT 1543
Qy 710 acatctacatgagacgcccaccatccgctggtcttggttggtcccaataatgcagcgat 769
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1544 ACAACTACATCGACGCGCGCCACACGAGGTGATCGGCTGGAGCTTCCAACTTCGGCCCT 1603
Qy 770 acgtacaggaagtcgaagaagtcctaaag---cgaacatcgggtgtcaagagcagcag 826
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1604 CCGGGACATCTTCTTACAGAGCGCGCAACGCTCCGCTCCACCGTGGACTACATCGACG 1663
Qy 827 gcttcgcacaaacagcgcaattacacgcccgttgaaagagcgtttatgacgccaacc 886
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1664 GCTTCACTCCACACGCGCAACTACTCGGCGCACTGTGGAGCCGTACTGGACGTCAACG 1723
Qy 887 agcaagtcggcgagacgcccgtgtgagtcggaaattctacacagttgaatcctgacatcg 946
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1724 GCACCGTTAAGCGCCAGCTATCCGCCAGTCCAGTGGGTGACTGGAACAGTACGTG 1783
Qy 947 acgaagcgcagctacggtgtgactgtacatccggtcgttcggtccggtcttccaagca 1006
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1784 ACGAGCTCTCTTGTCTACAGACCTGTGTCAGGCGCTGATGCCAAGGCTTCGGTCCG 1843
Qy 1007 gcatcgcatgctcatgacacacttacgcaacggttgaggtgtgcgaagcaaacacag 1066
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1844 ACATCGGTATGCTATGACACCTCCCGCAAGGCTGGGGTGGCCCAACCGTCCGACCG 1903
Qy 1067 gcccgagcaccgagaccagatgtcaacacctgtcgaacacagttcgaagattgacctggc 1126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 1904 GACGAGCTCTCCACGACGACCTCAACACTACGTTGACGAGAGAGCCGTATCCAGCCGCCGTA 1963
Qy 1127 agcaaccgagcctgtgtgacaaacgaagcgtlbgcgccctgagccagccgcgcaagca 1186
Db 1964 TCCACCCCGGTACGTGAGCAACGAGCCGGTGGCGGCTGGCCAGC----- 2011
Qy 1187 gccagacgacttccgacgagcgaacctcgacggtatgtctgtatgaagcgcgagtg 1246
Db 2012 GGGCCACGCTAACCCCGCTCCGGTGTGAGCGCTACGTCGTGGGTGAAGCCCGGGGTG 2071
Qy 1247 agtcgagcagcaagcgtctgagcgaatccgacaactgcaagaagtcgagcccatgt 1306
Db 2072 AGTCGACGCGGCCGACGAGAGATCCGAAAGAGAGGCAAGGCGCTTCGACGCGCATGT 2131
Qy 1307 ggcagccgactgaacgagcgtcgtaagcggtacgtacgaac-----gcgtacga 1357
Db 2132 GCGACCCGACCTCAAGGAGGCAACGCCGCCCAACGCGACACCCCTCGGGTGGCGCCA 2191
Qy 1358 actcccgatcgccgagcagtggttccgagcagtttgacagagttgtcggaagcgaac 1417
Db 2192 AGGCCCCATCTCCGGCCACTGGTCTGTGCCCACTCCGAGAGCTGCTGGCCACAGCCT 2251
Qy 1418 ggcagcagtgccgagcagtcagcagcagtcagccgcgctccgcgcgagtcgcgtcg 1477
Db 2252 ACCCGCCTCTGMAAGCGAGGAGGCAACGCGCTGACAGCGCTCAACGAGAACTGATCAG 2311
Qy 1478 ctgcgcgagtcgagccgagtcgagccgagccgagcagtcgcatcgccg 1527
Db 2312 CACCTCTTACGCGGAGAGCGGCGCTGCTCCCTCGGGGCGGCGCG 2361

RESULT 4

US-08-265-429A-4

Sequence 4, Application US/08265429A

Patent No. 5677151

GENERAL INFORMATION:

APPLICANT: Wilson, David B.

APPLICANT: Walker, Larry P.

APPLICANT: Zhang, Sheng

TITLE OF INVENTION: Thermostable Cellulase From A Thermomonospora Gene

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESS: Hodgson, Russ, Andrews, Woods & Goodyear

STREET: 1800 One M&T Plaza

CITY: Buffalo

STATE: New York

COUNTRY: United States

ZIP: 14203-2391

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage

OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1

SOFTWARE: Wordperfect for Windows 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/265,429A

FILING DATE: 24/06/94

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Nelson, M. Bud

REGISTRATION NUMBER: 35,300

REFERENCE/DOCKET NUMBER: 18617.0008

TELECOMMUNICATION INFORMATION:

TELEPHONE: (716)856-4000

TELEFAX: 716-849-0349

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1269 nucleotides

TYPE: nucleic acid

STRANDEDNESS: double-stranded

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

ORIGINAL SOURCE:

ORGANISM: Thermomonospora fusca

STRAIN: YX36
CELL TYPE: bacterium
US-08-265-429A-4

Query Match 8.5%; Score 314.6; DN 1; Length 1269;
Best Local Similarity 57.6%; Pred. NO. 4.7e-52;
Matches 732; Conservative 0; Mismatches 479; Indels 60; Gaps 7;

Qy 200 tgcataaccgctactgagcgaagagtagagagcgaagcgaacgaacaaatgcca 259
Db 17 TCGACAAACCCGTTGAGAGGGCCGCAAGCGTACGTACAAACCCGTTGAGAGGGCCG 76
Qy 260 ctctcgacgcaaaatgcgctgcttccacatctcgaagcgctgtagatgagca 319
Db 77 CCGCTGACCCGGCGGTTCGCGGTCGCAAGAGATCCACCGCTGCTGCGGAGCCGTA 136
Qy 320 tgcgtcgatcaagcgctcaacgagcgaacgaggttagagcatatctgagcgccg 379
Db 137 TCGGCGCCATCGAGGGCAACGACACCCGACACCGCTCCATGCGGTGCGCGACACC 196
Qy 380 t-----ctccagagcagaggaacacccctgaatcatltagatltgcatctag 430
Db 197 TCGAGAGAGCGCTCCGCCAGTCCGGTGGGAGCCGCTGACCATCCAGGTGCTCATCTACA 256
Qy 431 atctgcgagcagcagcagtcgagcgtcgcctccacagcggaactgcccgtacgag 490
Db 257 ACCTCGCGCGCGCGACACGCGCGCGCGCTGCTCCACAGGTGAGCTG-----GGTCCG 310
Qy 491 caggttgcagacctaataaagcgtatcatgcatgcatgcatgcatgcatgcatgcatg 547
Db 311 ATGAACCTGACCGCTTACAGAGCGAGTACGACCCGATCCGCAACATCATGTGGACT 370
Qy 548 atccgaagctactcagcagcagtcgagcagcagcagcagcagcagcagcagcagc 607
Db 371 TCGGAGACTACGAGAACTCGCGGATCGTCCCATATGAGATGACATCCTCCGCCAAC 430
Qy 608 cgtgacccaatctgagcattcaagcgtgtgcaagcggtgctgcat----- 654
Db 431 TCGTCAACCACTGGCGCGGAGGCGCGGACCGGACCGACTGCTGCTCACTCACTCACTCA 490
Qy 655 -----tacgagagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 709
Db 491 GCGGCTACGTCACAGCGGTGTGCTGACGCTCCGCAAGCTGGGAGATCCGAGAGCT 550
Qy 710 acatctacatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 769
Db 551 ACAACTACATGAGCGCGCGGACCAACGCGGTGATGCGGTGGGACATCACTTCCGCCCT 610
Qy 770 acgtacaggaagtcgaagagtcctcaagc--cgagcatcggtgtcaagcagcagc 826
Db 611 CGGTGACATCTTACAGAGGCGCGGACCGCTCCGCTCCACCTGAGACTGAGTCACTG 670
Qy 827 gcttgacacaaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 886
Db 671 GCTTATCTCCAAACGCGGCAACTCTGCGGCTGAGAGGCTGAGTCACTGAGTCACTG 730
Qy 887 agcaggtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 946
Db 731 GCACGTTAAAGCGCGGACGATCCGCAAGTCAAGTGGGTTCAGGAAACGATGACGTG 790
Qy 947 acgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1006
Db 791 ACGACCTCTCTTCTCCAGGACCTGCGTCAAGGCGCTGATCCCAAGGAGCTTCCGCTCG 850
Qy 1007 gcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 1066
Db 851 ACATGGATGCTCATGCAACACTCCCGCAAGCGGTGGGAGGCGGAGCGTCCGACCG 910
Qy 1067 gcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1126
Db 911 GACGAGGCTCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 970


```

QY 1127 agcaccgagcctgtgtgtcaaccagaacggtgtggtgctcgtgacccgacgagca 1186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 971 TCACCCCGCTACTGTGTCAACAGCGCGGCTGCGGCGCTGCGGCGAGCGGCC- 1023
QY 1187 gcccgagcagcttcgcgaacgcgacacctgcagcgcgfatgtctgtatcaaggccgcgggtg 1246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1024 -----ACGGTCAACCCCGCTCCCGGTGTGTGACCGCTACGTGTGTGTAAGCCCCGGGTG 1078
QY 1247 agtcgagcagcaacagcgtctgcagcgtacgcgaactgtgcgaagaagtcgcgaacatgt 1306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1079 ATCTCGACGCGCGCAGCAGAGATCCGACACGACGAGGCGCAAGGCTTCGACCGCATGT 1138
QY 1307 ggcagcgcagctacacgagctgtacggtgtactgaccaa-----ggttaccga 1357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1139 GCGACCGGACCTACACGAGGCAACGCCCGCAACGCAACAAACCCCTCGGTCGCTGCCCA 1198
QY 1358 atctcccgatgcgcgcgcagtggttcgcgcgcagtttgacacgctgtgtcgcgaagca 1417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1199 AGCGCCCACTCTCGGCGCACTGTCTGTGCCCCAGTTCCGCGAGCTGTGCCCAACGCTT 1258
QY 1418 ggcagcagtg 1428
    || | ||
DB 1259 ACCCGCTCTG 1269

```

```

RESULT 5
PCT-US95-09069-4
: Sequence 4, Application PC/TUS9509069
: GENERAL INFORMATION:
: APPLICANT: Wilson, David B.
: APPLICANT: Walker, Larry P.
: APPLICANT: Zhang, Sheng
: TITLE OF INVENTION: Thermostable Cellulase From A Thermomonospora Gene
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
: STREET: 1800 One Mt Plaza
: CITY: Buffalo
: STATE: New York
: COUNTRY: United States
: ZIP: 14203-2391
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
: OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
: SOFTWARE: Wordperfect for Windows 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/09069
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: U.S. Serial No. 08/265,429
: FILING DATE: 24/06/94
: ATTORNEY/AGENT INFORMATION:
: NAME: Nelson, M. Bud
: REGISTRATION NUMBER: 35,300
: REFERENCE/DOCKET NUMBER: 18617, 0008
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (716)856-4000
: TELEFAX: 716-849-0349
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1269 nucleotides
: TYPE: nucleic acid
: STRANDEDNESS: double-stranded
: TOPOLOGY: linear
: MOLECULE TYPE: genomic DNA
: ORIGINAL SOURCE:
: ORGANISM: Thermomonospora fusca
: STRAIN: YX36
: CELL TYPE: bacterium
: PCT-US95-09069-4

```

```

Query Match      8.5%; Score 314.6; DB 5; Length 1269;
Best Local Similarity 57.6%; Pred. No. 4,7e-52;
Matches 732; Conservative 0; Mismatches 479; Indels 60; Gaps 7;

QY 200 tcgtcaaccctctctgtggtgcgaagaagttacagagcgaagcgcgcaaccagaatgcga 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17 TCACCAACCCCTTCGAGGGGCCCAAGCTGTACGTGAAACCCGCTGTGCTGCGCAAGGCCG 76
QY 260 ctctgcagcgaanaatgacggtctgttcacatattcgaagcgctgtgtatgtgacgcga 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77 CCCTGAGCCGGCGGCTTCGCGGCTGCGCAACGATCCACCGCTGTGCTGCGCAAGGCCG 136
QY 320 tcgctgcatacgaacgcgttcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 137 TCGGCGCATCGAGGCGCAACGACAGCCCGCACACACCGGCTCATAGGCTCTGCGGACAC 196
QY 380 t-----ctccagcagcaggaacacacccctgaatcattgatatgtcatcag 430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197 TGGAGAGGCGCTTCGCGCACTCGGTGCGACCGGCTGACCATTCACGCTGTCATCTACA 256
QY 431 atctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 257 ACCTGCCGCGCGCGACTGCGCGCGCTGCTTCACAGGTGAGCTG-----GCTCCG 310
QY 491 caggttgcagacctatgaacgcagttacatccgattgcagatcctgag---ca 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 311 ATAACTCGACCGCTACAAAGAGATACATGACCGGATCGCGACATATGTGGGACT 370
QY 548 atccgaagttcccaagcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 371 TCACGACACTCGAGAACTCTCGGATCGTCCCATCATCGAGATCGATCTCCCTGCCCA 430
QY 608 cgttcaccaatagagcattcaagcgtgtgtgcaagcgcgcgcgcgcgcgcgcgcgcgcgc 654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 431 TCCTCAACCAAGTGGGCGGAGAGCGCGCACCGACTCTCGCTTACATGAACGAGACG 490
QY 655 -----tacgaagcagatcagtaacgcgtcgaacgaatgcgcgcgcgcgcgcgcgcgc 709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 491 GCGGCTACGTCACAGGTGTGGGTACCGCTCCGCAAGCTGGGCGAGATCCGACACTCT 550
QY 710 acactacatgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 769
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 551 ACAACTACATCGAGCGCGCCACGAGTGTGATCGGTGGACTGCCAATCTCGGCCCT 610
QY 770 agctacaggaatccgaaggtcctaag---cgagatcgcgggtgtcaagcgcgcgcgc 826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 611 CGGTGACATCTTCTACGAGGCGCGCAACGCTCCGCTCCACCTGACCTACGTGCACG 670
QY 827 gcttcgtaacaaacagcggaattacacgcgcgttggaagcgcgttcaatgaacgcgcgcgc 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 671 GCTTCATCTCAACACGCGCACTACCTCGGCACTGTGGAGCGGTACTTGAGCTCAACG 730
QY 887 agcaagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 731 GCACCGTTAACGGCGCAGCTCATCCGCCAGTCCAAAGTGGTTGACTGGAACAGTACGTCG 790
QY 947 acgaagcgcgaactacgcggtgtgaattgtactcgcgcgcgcgcgcgcgcgcgcgcgcgc 1006
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 791 ACAGAGTCTCTTCGTCACAGGACTGTGTCAGGCCCTGATGCGCAAGGCTTCGCGTCC 850
QY 1007 gcatcgcatgtcatcgaacacttaacgaacggtgtggtgtgttcgaagaacaaacga 1066
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 851 ACATCGGTATGCTCATGACACTCCCGCAACGGCTGGGTGGGCCCAACCGTCCGACCG 910
QY 1067 gcccgagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 911 GACCGAGCTCTCTACCGAGCTCAACACCTACAGTTCAGAGACCGGATGACCGCGCTA 970
QY 1127 agcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 971 TCACCCCGCTAACTGTGTCAACAGCGCGGCTGCGGCGCTGCGGCGAGCGGCC----- 1023

```

QY	1187	gcccagagagctccccaagcgccacctgagcagctatgtctgatacaagccgcgggtg	1246
Db	1024	-----AGGTCACGCCGCTCCGGGTGTGACCCCTACTCTGGGTGAAGCCCCGGGTG	1078
QY	1247	agtcgagcagcacaagcgcgttgagcagatccgacaactgycagaagaagtcgagcccatgt	1306
Db	1079	AGTCCGACGGCGCGCAGCAGAGAAATCCGACACGACGAGGCGCAAGGGCTTCGACCGCATGT	1138
QY	1307	gcgacccgagctaacgaacgtctgtaggggtactgacaaac-----ggattacga	1357
Db	1139	GGGACCCGACTACCAAGGGAACGCCCGCAAGGGCAACACCCTCTGGGTGGCTGGCCCA	1198
QY	1358	actcccgatctgcgcgcagctggtctccgcgcgcagtttgacagcttgttcgcgaagcgcac	1417
Db	1199	ACGGCCCAATCTCGGCGCACTGTCTCTGCGCAAGTTCGCGAGCTGCTGGCCAAACGCT	1258
QY	1418	ggccgagcggtg	1428
Db	1259	ACCCGCCCTGTG	1269

```

RESULT 6
US-09-136-574A-1
: Sequence 1, Application US/09136574A
: Patent No. 6294366
:
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11707 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-136-574A-1

```

```

Query Match Similarity      3.6%; Score 134.2; DB 4; Length 11707;
Best Local Similarity      52.6%; Pred. No. 2,2e-17;
Matches 392; Conservative  0; Mismatches 338; Indels 15; Gaps 4

OY 1559 gcgcagccgcagtcgcagccagccagtcagtcgcgcgcgtccgcagctcgcagccgct 1618
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3835 CACCAGACACCGGCACCGACCCGACCCGACATCGACGCGCACACCAACCACTTACACACCTG 3894
OY 1619 ctccgtgcgcagccgcagtcgcgaacccgcagtcagctgcgcgtcgcgcgtccgcagctcga 1678
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3835 CACGCACATCAGCCCCGACACCGACCCCAACAGTACAGCAACAGCGCGCATCTCAACGCCGA 3954
OY 1679 gccgccttcgcgtccgcagccgcagtcgcgcagccagtcagtcgcgcgtcgcgcgtccga 1738
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3955 CCGCGACAGTGCAGGGTTACTGTGACTCGCACACCGACACCAACACCGACGCGCACCGCA 4014
OY 1739 cgtgcgcgcggtgcgcg---gtggcgtgaagtggtgcaglaagaacaatgaltcgcgcg 1795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4015 CAGGCACACCTGCGACCGGCAAGTGGTTGCAAGCTACTATACCAAGAAACAATGAGACAAGTG 4074
OY 1796 cgggtgataaccaagtcacaacccgggtctccagtttgtagataccgggtcgtcgtcgtg 1855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4075 CGAGCACAAGTCTCTTTAAGCCGCTGTTTAAATATGTGATGAGGACACACAGTGTG 4134
OY 1856 atttgcagcaggtgcaggttgcgtatcgttaccgcggatgtgtgtcgtcgcacatg 1915
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4135 ATCTTAGCAGGGGTTAAGATTAAGATCTGCTACACAGTGATGGTG---ACAAGCCACANA 4191
OY 1916 tgtlaaacctgtgactggcgcgcgatgggggtgtgtggaaataaccgcgcctcgtcgcgtcg 1975
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4192 GTGCGGTTGTGTACGTGGGCACAGATTAAGGGGCAAGCAATGTGACATTTTGTGAACG 4251
OY 1976 tgaacccgcgcagcgcgcagcgcgcgcacccacccacccagttgtcgtcagtcgtggaac-- 2033
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4252 TGAGCAGCGGAGTGAAGTGAGCGCGGATTTATTCTTGAGGTAGGATTTTGCATGGAGCTG 4311
OY 2034 ----gttgcgcgtgtgtgtgtgcgcagcgggttgagttcaaaaaccgggttgaaagtgact 2089
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4312 GGCAATTCGAGGCTGTGAAGACACAGGGGAAATACAGTTAAGTTTAAAGAAATGACT 4371
OY 2090 gttcgaatttccagagaccaaatactatctgtatgggagcgaacccacttccagact 2149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4372 GAGGCAATTTACATATAGCGACAGCACTGCTATGCTGTGACAGACATGACGAATTTATGAG 4431
OY 2150 gga---cgaagtgacggtgtacgtcaacgcgcgttggltgtggggagacgtacgcgtcgcg 2206
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4432 AGAATGCGAAGGTAAACGCTGTATGTATGATGCTGTTCTGTATGAGGGGCGAGCGGGCGG 4491
OY 2207 gcaacagcccccagcccccacacatcccccagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4432 GAGGCACACTTGCACCGACACGACACACGACACACCAACGCCAATCTCGACAGACAGCA 4551
OY 2267 atgtgacgcgcgcgcagtggtccgac 2291
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4552 CACGCAGCGCGCACGACCAAGCCCAAC 4576

RESULT      7
US-07-862-588-1
; Sequence 1, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Sch Jeln, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

```



```

; PRIOR APPLICATION NUMBER: PCT/US99/11971
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Nearest "neighbor" - Streptomyces
; OTHER INFORMATION: thermovibaceus
US-09-321-981-2

Query Match      2.6%; Score 95.2; DB 4; Length 1116;
Best Local Similarity 51.2%; Pred. No. 4.6e-10;
Matches 381; Conservative 0; Mismatches 343; Indels 20; Gaps 6;

QY 2611 accaggtgcaaggcgacgaataacccgggtgcaagcaatgagtggaatttcgtgcgccag 2670
DB 121 accaccagatccagacccggtacgtgtgtagaacaacacgcgtggcgacagcgccacc 180
QY 2671 cagtgctccatccatccatccacgcgacgtgctctgagacggtagagccactgcgaactcagc 2730
DB 181 cagtg-----catcaatgtgaccggcaacaggttctgagatacccccagggcgacgttcg 234
QY 2731 ggttggaaccggcggtgaccccgacgacgtatccgtatcacagggtctgcacgtgggc 2790
DB 235 gtgcgaccacacggcgcccggaagtctctacccctcgtctacgacggtctgcacacagc 294
QY 2791 aactgcacacgaagaagactcgaggtcgatccagatccagatcgaattgttgctgtgtg 2850
DB 295 aactgcgaccccgccgcagac---gctgcacatgctggtatccgtctgacgtgcagcgccc 331
QY 2851 acgtctgtgagtagacgaaggtgtctgtcggtcgatgacgtggtgcctacgacatttgg 2910
DB 352 agcagtgatctctacacgacacacgcgacggtctctacacgcgacccgacgtctcgtc 411
QY 2911 acgaacagtatacccaacgaacacgcgttcagccaacggtatccgaaatcatgatttgctg 2970
DB 412 ctggacccacacacccgcacacacgggtgtg---aaccggaacgagatcatgattgttc 468
QY 2971 aattcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3030
DB 469 aactgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 525
QY 3031 ggtcacacgtggaatgtctgtgacaggttcacgaacacgtctgtggaagatttatctcagtc 3090
DB 526 ggcgcgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 582
QY 3091 ctgaaccccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3150
DB 583 ctggcccccctccgcgaacacacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 642
QY 3151 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3210
DB 643 gtacagcaacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 702
QY 3211 atctggcaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3270
DB 703 ccgttggaagggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 761
QY 3271 agtctcagcccgacacacgacccgacccgacccgacccgacccgacccgacccgacccg 3330
DB 762 -gggcgggaacggcgacaccccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 820
QY 3331 tctcgaagcccgaccccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3354
DB 821 acacctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 844
```

```

RESULT 11
US-09-197-649-7
; Sequence 7, Application US/09197649
; Patent No. 6194550
; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Tuerk, Craig
; APPLICANT: Prihnow, David
; APPLICANT: Smith, Jonathan D.
; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
; FILE REFERENCE: NEX02/CI-CON
; CURRENT APPLICATION NUMBER: US/09/197,649
; EARLIER FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: 07/829,461
; EARLIER FILING DATE: 1992-01-31
; EARLIER APPLICATION NUMBER: 07/739,055
; EARLIER FILING DATE: 1991-08-01
; EARLIER APPLICATION NUMBER: 07/561,968
; EARLIER FILING DATE: 1990-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 7
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
; OTHER INFORMATION: fragments having NcoI restriction sites.
US-09-197-649-7

Query Match      2.6%; Score 94.8; DB 4; Length 390;
Best Local Similarity 55.0%; Pred. No. 4.8e-10;
Matches 186; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 1418 ggcacgaggtgcagacgtcagacacgacgacccgcgcgtccgcgcgacgtctgcgtg 1477
DB 2 ggcacatggaagacgaacacgacacgacacgacacgacacgacacgacacgacacgacacg 61
QY 1478 ctctgcgagtcgagccgacgagtcgagcccgacgacgacgtgcacatctgcgcgtctc 1537
DB 62 cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 121
QY 1538 cgagctcgaacccgctctcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1597
DB 122 cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 181
QY 1598 cgctccgagtcgagccgctctcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1657
DB 182 cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 241
QY 1658 ctctgcgcgtctcgaagtcgagccgctctcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1717
DB 242 cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 301
QY 1718 gctcgcgcgtcgcgtctcgaagtcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1755
DB 302 cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 339

RESULT 12
US-09-197-649-7/c
; Sequence 7, Application US/09197649
; Patent No. 6194550
; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Tuerk, Craig
; APPLICANT: Prihnow, David
; APPLICANT: Smith, Jonathan D.
; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
```

BEST PAGE AVAILABLE COPY

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☒ **FADED TEXT OR DRAWING**

☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

